THE UNIVERSITY *of* EDINBURGH

Elucidating the Early Events of Protein Aggregation using Biophysical Techniques

Harriet Lucy Cole

Ph.D.

The University of Edinburgh

2012

The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at and repair.

– Douglas Adams –

Declaration

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Edinburgh. Unless otherwise stated this work is my own and has not been submitted for any other degree or professional qualification.

Chapter 3: All simulations were performed by Jason Kalapothakis.

Chapter 4: All simulations were performed by Massimilliano Porrini.

……...

Harriet Lucy Cole

August 2012

Acknowledgements

I would firstly like to acknowledge and thank my supervisors Dr Perdita Barran and Professor Cait MacPhee for all the help, guidance and support they have given me since I was a senior honours student. Without their assistance none of this would have been possible.

A special thank you to all my fellow group members over the last four years for everything, especially helpful discussions, cakes, cups of tea, office singing, bodging of lab equipment, group meetings and advice concerning where the bit of hardware left over when the instrument was reassembled should have gone. So, in vaguely chronological order, thank you: Jason, Peter, Stefan, Roland, Martin, Jude, Yana, Ewa, Sophie, Kamila, Bex and Jakub. Thank you also Max for your invaluable help in the face of my luddite approach to molecular modelling. Finally I am grateful to my project students Tom and Callum for all their hard work.

I would like to acknowledge SIRCAMS members Stefan and Logan for their assistance with FT-ICR MS measurements. I am also grateful for all the technical help provided by Stuart and Davey from the mechanical workshop and Donald from the electronics workshop. A big thank you also to Jane Patterson and the fantastic teams in Chemistry Stores and in the Chemistry Administrative Office.

Finally I am indebted to my family, Chris and my friends for their love and support, cheering me up and providing a different perspective. A special mention to Ludo and Scrabble for sitting on my lap and not letting me get up until I had written something.

Abstract

Proteins and peptides can convert from their native form into insoluble highly ordered fibrillar aggregates, known as amyloid fibrils. The process of fibrillogenesis is implicated in the pathogenic mechanisms of many diseases and, although mature fibrils are well characterised by a plethora of biophysical techniques, the initiation and early steps remain, to date, ambiguous. Mass spectrometry can provide invaluable insights into these early events as it can identify the low populated and transient oligomeric species present in the lag phase by their mass to charge ratio. Recent evidence has shown that oligomers formed early in the aggregation process are cytotoxic and may additionally be central to the progression of diseases associated with amyloid fibril presence. The hybrid technique of ion mobility mass spectrometry can be employed to provide conformational details of monomeric and multimeric species present and elucidate the presence of oligomers which possess coincident mass to charge ratios. Molecular modelling, in conjunction with experimental results, can suggest probable monomeric and oligomeric structural arrangements.

In this thesis three aggregating systems are investigated: amyloidogenic transthyretin fragment (105-115), insulin and two Aβ peptides. Initially amyloidogenic endecapeptide transthyretin (105-115) is studied as it has been widely utilised as a model system for investigating amyloid formation due to its small size. Secondly insulin, a key hormone in metabolic processes, is investigated as extensive research has been carried out into its aggregation into amyloid fibrils. The formation of insulin amyloid fibrils rarely occurs *in vivo*; however localised amyloidosis at the site of injection and the aggregation of pharmaceutical insulin stocks present problems. Thirdly the aggregation of Aβ peptides Aβ(1-40) and Aβ(1-42) and their interactions with an aggregation inhibitor, RI-OR2, are characterised. Aβ(1-42), although less commonly produced *in vivo*, is more cytotoxic and has a faster aggregation mechanism than Aβ(1-40). Both Aβ peptides are implicated in the aetiology of Alzheimer's disease whilst RI-OR2 has been reported to prevent the production of high molecular weight oligomers, with particular suppression of Aβ(1-42) aggregation.

Abbreviations

Aβ	amyloid beta peptide
AFM	atomic force microscopy
APP	amyloid precursor protein
ATD	arrival time distribution
AUC	analytical ultracentrifugation
β_2m	β_2 microglobulin
CCS	collision cross section
CD	circular dichroism
CID	collision induced dissociation
Da	Dalton
DC	direct current
DT	drift tube
EHSS	exact hard sphere scattering
ESI	electrospray ionisation
EM	electron microscopy
FAC	familial amyloid cardiopathy
FAIMS	field asymmetric ion mobility spectrometry
FAP	familial amyloid polyneuropathy
FRET	Förster resonance energy transfer
FT-ICR	fourier transform ion cyclotron resonance
FTIR	fourier transform infrared spectroscopy
HCl	hydrochloric acid
HDX	hydrogen-deuterium exchange
IAPP	islet amyloid polypeptide
IM-MS	ion mobility mass spectrometry
IMS	ion mobility spectrometry
LS	long straight (of fibrils)
MCP	microchannel plate detectors
MD	molecular dynamics
MM-PBSA	molecular mechanic - poisson-boltzmann surface area

MoQTOF	mobility quadrupole time-of-flight
MS	mass spectrometry
m/z	mass to charge ratio
NaI	sodium iodide
NMR	nuclear magnetic resonance
n-ESI	nano-electrospray ionisation
PA	projection approximation
pdb	protein data bank
PME	particle mesh ewald
PrP^{C}	prion protein: cellular form
PrP^{Sc}	prion protein: scrapie form
QTOF	quadrupole time of flight
RF	radio frequency
R_g	radius of gyration
REMD	replica exchange molecular dynamics
RMSD	root mean squared deviation
SA	simulated annealing
SEC	size exclusion chromatography
SSA	senile systemic amyloidosis
SSNMR	solid state nuclear magnetic resonance
SRCD	synchrotron radiation circular dichroism
TEM	transmission electron microscopy
TFE	trifluoroethanol
TIC	total ion count
TM	trajectory method
TOF	time of flight
Trp	tryptophan
TSEs	transmissible spongiform encephalopathies
TTR	transthyretin
TWIMS	travelling wave ion mobility spectrometry
TWIG	travelling wave ion guide
VMD	Visual Molecular Dynamics programme

WL	worm like (of fibrils)
WT	wild type

Physical Constants and Quantities

d	distance
e	elementary charge, 1 e = 1.6021765 x 10^{-19} C
E	electric field strength
E_k	kinetic energy
K	mobility
k_b	Boltzmann constant, $k_b = 1.3806488(13)\times10^{-23}$ JK^{-1}
K_d	dissociation constant
m	mass
m/z	mass to charge ratio
n	oligomeric order
N	number density
P	pressure
q	net charge
t	time
T	temperature
Td	Townsend unit, 1 Td = 10^{-17} Vcm^2
v	velocity
V	potential difference
z	nominal charge
μ	reduced mass
ρ	density
ω	angular frequency
Ω	momentum transfer intergral

Table of Contents

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Appendices

Table of Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

1

Introduction and Motivation

The ability of proteins and peptides to convert from their soluble native form to insoluble highly ordered fibrillar aggregates is the subject of intense scientific interest and relevance. The formation and presence of these structures, known as amyloid fibrils, is implicated in the development of degenerative diseases. The precise processes by which this conversion occurs are still elusive, particularly in their initial stages, but mass spectrometry, ion mobility mass spectrometry and molecular modelling can provide invaluable insights into the early events which occur in the fibrillogenesis pathway.

1.1. Introduction

The first observations of amyloid plaques in the organs of the deceased date as far back as 1639 [1]. Amyloid plaques were so named by Rudolph Virchow, who in 1854 discovered the positive iodine staining of cerebral plaques, and thus identified them as starch, from which the term amyloid originates (derived from the Latin 'amylum' and Greek 'amylon'). Five years later Friedreich and Kekule refuted the supposition that amyloid plaques were composed of carbohydrates, correctly identifying their proteinaceous nature based on their high nitrogen composition. Latterly researchers have discovered over twenty proteins (Table 1.1) which can convert from their soluble native state into insoluble fibrillar aggregates *in vivo* [2]. These proteins vary hugely in their primary sequence, tertiary structure and evolutionary function, however, post fibrillogenesis, the aggregates can be of a startlingly homogeneous fibrillar structure composed predominantly of a single protein or peptide [3,4]. The histological diagnostic test for amyloid fibril presence is the chemical binding of Thioflavin T [5] and green birefringence observable on binding Congo Red dye [6]. The biophysical definition of an amyloid fibril, whether *in vivo* or *in vitro*, is one that possesses "cross-β" structure where the β-sheet fibril core is perpendicular to the fibril axis [7].

The disparity of amyloid-forming proteins and peptides, coupled with the homogeneous and highly stable fibrils produced, suggests that cross-β-sheet structure may be a global potential energy minimum for many proteins, energetically preferential to their native state [8]. This assertion is supported by the findings that, under denaturing conditions, proteins unconnected with amyloid disease can form amyloid fibrils *in vitro* [9]. It is postulated that amyloid fibril formation is a generic property of all polypeptide chains as the hydrogen bonds which stabilise the fibril core structure involve the peptide backbone which is shared by all proteins [7,10].

1.2. Aggregation *in vivo*

Fibrils are stable under physiological conditions because of their numerous strong hydrogen bonds. A protein's external environment is already carefully homeostatically regulated, particularly in terms of pH and temperature, but in addition, to prevent fibril formation, proteins are chaperoned during folding transitions [9]. Misfolded or unfolded proteins are targeted and destroyed by chaperones and housekeeping mechanisms before they can form fibrils [11]. However when the homeostatic control or the degradation mechanisms which deal with misfolded and unfolded proteins fail, pathogenesis can occur [12].

A diverse range of diseases, known as amyloidoses, are characterised by amyloid fibril presence. A decreased efficacy of the cellular response pathways which deal with unfolded or misfolded proteins, combined with increased propensity for protein damage to occur with age, could explain why the majority of amyloidoses develop late in life [9]. In the developed world, with its aging population, these are predicted to become increasingly prevalent. Some amyloidoses and their associated proteins are shown in Table 1.1.

Disease	Associated Protein(s)
Alzheimer's	Aβ peptides and Tau protein
Atherosclerosis	Apolipoprotein A1
Dialysis-related amyloidosis	β_2-microglobulin
Familial amyloidotic polyneuropathy	Transthyretin
Familial non-neuropathic amyloidosis	Lysozyme
Huntington's disease	Huntingtin
Injection-localized amyloidosis	Insulin
Parkinson's syndrome	α-synuclein and Tau protein
Primary systemic amyloidosis	Immunoglobulin light chain
Senile systemic amyloidosis	Transthyretin

Transmissible spongiform encephalopathies	Prions
Type-II diabetes	Islet amyloid polypeptide

Table 1.1 | Examples of amyloidoses and their associated proteins or peptides [10,13].

Amyloid diseases are classified into neurodegenerative conditions, non-neuropathic localised amyloidoses and systemic amyloidoses depending on the location of the amyloid deposits, known as 'amyloid plaques' [14]. In systemic and localised amyloidoses large quantities of fibrils build up, disrupting organ function [15]. However, for neurodegenerative conditions, studies have shown that pathogenesis can occur before the presence of mature fibrils [16] and there is no direct correlation between the quantity of fibrillar deposits and pathogenic effects [17]. It is suggested that aggregate plaques may be the end product of a process in which the principle disease-causing species are small, soluble oligomers that form early in the aggregation process [18]. Increasing numbers of studies have shown that preamyloidogenic forms of proteins exhibit cytotoxicity [17]; cytotoxic oligomers have been found in the early aggregation stages of α-synuclein [19,20], Aβ(1-42) [21,22] and transthyretin [23,24]. Whilst studies of the peptides implicated in the development of Alzheimer's disease have indicated that intermediate metastable states, not mature fibrils, could be responsible for pathogenesis [25,6].

Evidence has been collected which implies that the cytotoxicity of prefibrillar aggregates and soluble oligomers is not just confined to the proteins associated with known amyloid diseases [26]. Specific antibodies have been developed which are claimed to cross-react with early aggregates of different proteins, but not with their monomeric or fibrillar states, to inhibit their toxicity [27]. These studies suggest that the cytotoxicity of soluble oligomers which precede mature fibrils may have a common structure and pathogenic mechanism, and the formation of these may also be a common property of all proteins [27]. Most recently the discovery of a highly cytotoxic prion monomer with α-helical structure has broadened the scope of understanding

beyond oligomers and mature β-sheet aggregates [28]. If certain monomers and soluble oligomers are cytotoxic it is possible that amyloid fibril formation is an evolutionary protective mechanism to sequester these dangerous oligomeric species into a stable and harmless form.

Thus the characterisation of all of species, and conformers of species, present in the aggregation process is vital due to the role they may play in pathogenesis. The aggregation pathway is complex and can contain species which do not culminate in mature amyloid fibrils; these are known as 'off-pathway' products, whereas those which form amyloid fibrils are 'on-pathway'. Many of these species are non-covalently bound, transient and present in small quantities, making them difficult to observe experimentally, unlike amyloid fibrils which have a stable and ordered structure.

Perhaps unsurprisingly evolution has utilised the ability of proteins to form amyloid fibrils. These are known as 'functional amyloid proteins' due to the beneficial effect they confer upon their host. Functional amyloid proteins are utilised by bacteria, fungi, insects and mammals in diverse roles [3]. Some examples are the *Escherichia coli* curlin protein which assembles into fibrils to aid bacterial colonisation and binding with host cells [29] and, in humans, the formation of fibrillar striations along which melanin granules grow [30].

1.3. Aggregation *in vitro*

Globular proteins can be converted into amyloid fibrils *via* exposure to mild denaturing conditions which cause partial protein unfolding [31]. This allows the aggregation process to be studied *in vitro* and denaturing conditions include elevated temperatures [32,33], low pH [34], and the presence of organic solvents [35,36]. The ability of amyloid fibrils to self assemble, in response to changes in external environment, into a predictable ordered structure, in addition to intrinsic properties such as their high tensile strength, opens up a wealth of opportunities for biomaterial design [37]. One

example of this is the use of self assembling peptides to support tissue repair. Amyloid filaments are of similar dimensions to the biological scaffolds found in nature, but with the additional benefits of controlled assembly and disassembly and the potential to incorporate growth factors or drugs, allowing either slow drug release or drug delivery to specific sites [38]. A recent study showed that a self-assembled peptide network could support axon regeneration and restore vision in hamsters [39].

1.3.1. Structure of Amyloid Fibrils

A plethora of techniques have been used to study the structure of amyloid fibrils and their protofilament subunits. Until recently it was not possible to characterise fibrils at the molecular level due to their large size and non-crystalline structure, as these attributes create difficulties in probing their structure by either Nuclear Magnetic Resonance Spectroscopy (NMR) or X-ray crystallography. Therefore structural and morphological information was obtained from low resolution techniques such Transmission Electron Microscopy (TEM), Atomic Force Microscopy (AFM) and X-ray fibre diffraction [7,40,41,42,43]. Recently, the development of solid state NMR (SSNMR) [44,45,46] and achievements in growing small peptide crystals, which can be analysed by X-ray crystallography [47,48], have allowed amyloid fibrils to be characterised at much higher resolution. SSNMR can measure inter-nuclear distances and torsion angles of peptide systems [44,49], leading to the significant finding that fibrils possess a degree of almost crystalline uniformity, explaining, in part, their great stability. Other techniques which can be used to elucidate amyloid fibril structure are Circular and Linear Dichroism and Synchrotron Radiation Circular Dichroism (SRCD) which can provide detailed information on the secondary structure present in fibrils, including quantity of β-sheet, orientation of strands and inter- and intra-strand interactions. There have been recent exciting developments in fluorophore photophysics which have applications for self-assembling nanomaterials [50] and Förster resonance energy transfer (FRET) which allows the classification of different fibril structures [51,52]. Hydrogen Deuterium Exchange (HDX) can be used in conjunction with nano-electrospray ionisation mass

spectrometry to determine the tertiary and quaternary structure of fibrils by studying their solvent accessibility [53]. HDX, NMR and mass spectrometry have observed dynamic equilibria between fibrils and oligomers, owing to monomer/oligomer addition to fibrils and fibril fragmentation and re-association [54].

Typically amyloid fibrils consist of between two and six protofilaments which have a diameter of 2 to 5 nm [40]. Studies using X-ray fibre diffraction report that protofilaments possess cross-β structure with β-strands perpendicular to, and hydrogen bonds parallel to, the fibril axis [7,55]. There appear to be two categories of protofilament association; either a lateral arrangement which results in a long ribbon fibril of a width of up to 30 nm and thickness between 2 to 5 nm [56], or a twisted rope of protofibrils between 7 and 13 nm wide [40]. Most fibrils share a core region of between two or four β-sheets, however the length of these and whether they are parallel or antiparallel is variable [3]. SSNMR and X-ray crystallography techniques are in agreement with cryo-EM data which shows amyloid fibrils consists of pairs of flat β-sheets [57].

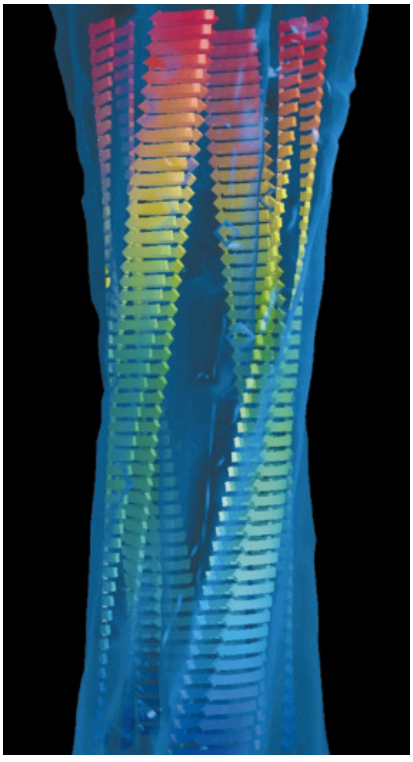


Figure 1.1 | Molecular model of a SH3 domain amyloid fibril derived from cryo-EM data. Four protofilaments twist together to form the mature fibril. The cross-β sheets are shown inside the EM density data, which is represented by the transparent blue surface. Figure and caption adapted from Dobson [10].

In addition to the heterogeneity which exists between amyloid fibrils composed of different subunits, fibrils formed by a single protein or peptide can exhibit varying architectures [56,57,58]. These are believed to originate from differences in the position of the polypeptide chains within mature fibrils [59].

Preformed seeds can propagate their molecular architecture, even in non favourable conditions, templating the resulting fibrils in their image [59]. This can play a key role in cross species prion transmission [60]. Fibrils can be formed with distinct morphologies, and the most kinetically favoured conformation, in a particular set of conditions, will be the most dominant. This is in contrast with that of a natively folded protein where one conformation is significantly more stable and exists at a well defined energy minimum [3].

Therefore, although properties of the polypeptide chain determine the framework of fibrillar architecture and the core region, variations in the sequences of different proteins and peptides lead to both differences in the diameter and the number of protofilaments which constitute a fibril and heterogeneous fibril morphologies. This is due to the unique side chain interactions of each protein or peptide sequence with each other and the solvent [3].

1.3.2. Fibrillogenesis

Fluorescence measurements have played a major role in monitoring fibril growth kinetics as fibrils bind Thioflavin T and Congo Red dyes, whereas prefibrillar forms and oligomeric species with no β-sheet structure do not [61]. The time course of fibrillogenesis generally displays sigmoidal kinetics, consisting of an initial lag phase followed by a period of accelerated growth [62,63]. These kinetics are believed to be the result of a complex combination of primary and secondary nucleation processes. Primary nucleation is the creation of a species which can elongate by monomer or oligomer addition, whereas secondary processes involve nucleation at interfaces and fibril fragmentation and branching. The length of the lag phase is very sensitive to environmental conditions such as pH, agitation, protein concentration [64] and the presence of salt [65]. Fibril formation can be accelerated by 'seeding' the solution i.e. the addition of preformed aggregate nuclei or oligomers. In the case of disease, seeding can accelerate the illness progression and even act as an "infectious catalyst." Such seeding could come about from the ingestion of an intermediate, as

exemplified by transmissible spongiform encephalopathies which can be spread *via* prions [66].

1.3.2.1. Transition from Native Structure

The conformational change required for the transformation of a protein's native structure to the cross-β-sheet form of amyloid fibrils, is often dramatic, and requires a partial or complete unfolding of the structure to occur [10,31,67]. This can be induced by a destabilisation of the protein's native fold by environmental conditions *in vitro* such as low pH, high temperatures or the presence of organic solvents [34,68,69,70]. Inherited protein mutations *in vivo* can destabilise a protein's native structure, leading to amyloid formation. There is a strong correlation between decreased conformational stability and aggregation propensity in disease related proteins [68,71]. Thus small molecules or ligands which bind to a protein and stabilise its native fold, thereby decreasing the likelihood of aggregation, can be used as therapeutics [2,72,73].

Despite the common propensity of proteins to aggregate into amyloid fibrils, the individual amino acid sequence of each peptide or protein has a profound effect on how probable this event is. Charge on the molecule is crucial, as highly charged species are prevented from self association by their mutual repulsion [36,74]. This is verified by the observation that if macromolecules are present in the solution, negating the high charge on the aggregating protein, aggregation will occur at a swifter pace [75,76]. Another key factor in the predisposition of a polypeptide chain to aggregate is the hydrophobicity of its residues. Once a protein's native fold has been disrupted, the exposure of previously protected hydrophobic regions can lead to aggregation. Evolution has selected against sequences of amino acids which promote the formation of β-sheets, such as those with many adjacent hydrophobic residues [77] or an alternating pattern of hydrophilic and hydrophobic residues [78], which would encourage aggregation into fibrils. Computer algorithms can predict the sequence sections of an unfolded polypeptide chain which are vulnerable to aggregation [79]; more recently this method has been extended to include folded proteins and those

with some degree of structure [80]. These calculations are particularly important as they enable the prediction of how mutational changes would effect the aggregation of proteins and may eventually aid in combating amyloid disease.

1.3.2.2. Prefibrillar Intermediates

In vitro a range of structures can be observed prior to the presence of mature fibrils: soluble oligomers, amorphous larger aggregates and diverse protofibrils. There is some confusion in the literature regarding the distinction between oligomers, protofibrils and protofilaments. For the purposes of this thesis oligomers are defined as clusters of non-covalently bound protein or peptide monomers, whilst protofibrils are aggregates of about 2 to 5 nm in diameter with some β-sheet structure and protofilaments are the components of fibrils [14].

Oligomers are observed very early in the lag phase (Figure 1.2a), where they exist in rapid equilibrium with the monomeric population. Protofibrils are structurally diverse and can be spherical or chain-like (Figure 1.2b). They are observed for a variety of amyloidogenic proteins such as transthyretin [81], β_2 microglobulin [70] and α-synuclein [19]. Their growth and structure is dependent on solution conditions and can be altered by changing pH, temperature or ionic strength [82].

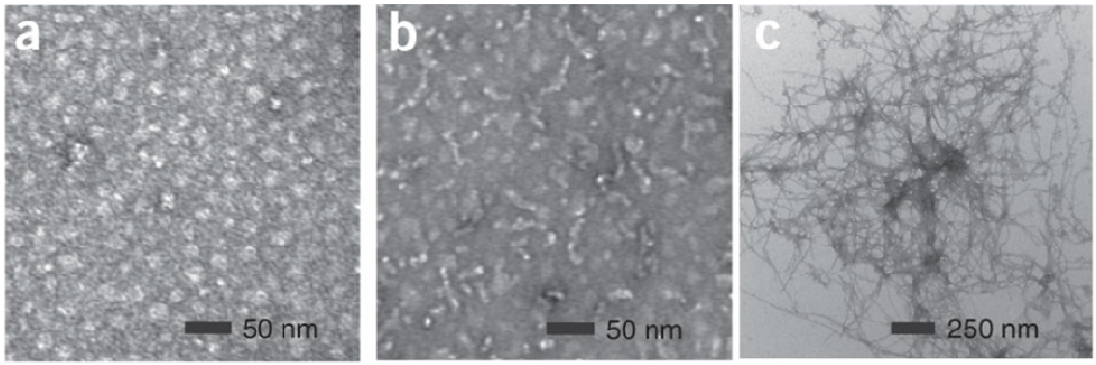


Figure 1.2 | Prefibrillar Aβ(1-42) intermediates (a) TEM of Aβ(1-42) oligomers incubated at 4 °C for six hours. (b) TEM of Aβ(1-42) protofibrils incubated at

*37 °C for six hours. (**c**) TEM of Aβ(1-42) fibrils incubated at 37 °C for 12 days. Figure and caption from Ahmed et al. [83].*

Protofibrils form from the assembly and β-sheet structural reorganisation of these small structurally disorganised oligomers [84,85,86,87,88,89]. The precise mechanism of protofibril growth is unknown as it is challenging to monitor this dynamic situation; it may be that a specific oligomeric species is responsible for nucleating protofibril formation, following which protofibrils grow by monomer or oligomer addition and/or the coalescing of protofibrils [88,90,91]. Protofibrils contain sufficient β-sheet structure to allow binding with Thioflavin T and Congo Red dyes, hence their detection. Protofibrils lengthen during the lag phase and can persist even when mature fibrils have developed, although their population is reduced [41,82,90]. Experimental data have shown that in some cases protofibrils are involved in the formation of fibrils [41,63], but conflicting data find they are off-pathway species [70].

Kayed *et al.* discovered the specific binding of certain antibodies to oligomers (and/or protofibrils, these are not distinguished by this study) of different proteins, but not with their monomeric or fibrillar states [27]. This suggests that there are structural elements common to these early order species, regardless of sequence variation. In large quantities these antibodies also mitigated the cytotoxic effects of oligomers (and/or protofibrils). Caughey and Lansbury have suggested that the cytotoxicity of oligomers/protofibrils (undistinguished in this study) comes from their ability to form unregulated membrane pores [17].

1.3.2.3. Fibril Assemblies

Mature amyloid fibrils can self associate to form assemblies, the most common of which are the visible amyloid plaques found in organs: the hallmark of amyloid disease. Other assembly structures which have been observed are spherulites [92,93,94], monodisperse particulate structures [95] and the lateral association of mature fibrils to form broad ribbons [57]. Spherulites, observed *in vivo* and *in vitro*, are tens of microns

in diameter and composed of radially orientated amyloid fibrils [92], in some cases with an amorphous core. They can be identified under cross-polarised optical microscopy by their Maltese cross extinction pattern [92] and are slow forming under conditions of high molecular charge. Particulate structures conversely are observed only *in vitro* when there is little or no net charge on the molecule [94] and have been predominantly studied in relation to whey products and food production. These structures tend to be spherical and hundreds of nanometres in diameter [95] and are composed of short, randomly arranged, amyloid fibrils.

1.3.3. Why Mass Spectrometry?

Mass spectrometry (MS) has a unique role to play in the study of the early events of fibrillogenesis, about which little is known. This is because the small, transiently populated heterogeneous oligomeric states which precede fibril growth, and could be the pathogenic agents in many diseases, are difficult to investigate using traditional biophysical techniques. To date oligomers have been characterised mostly using low resolution techniques such as size exclusion chromatography (SEC) [96], analytical ultracentrifugation (AUC) [97] and light scattering techniques [98]. SEC and AUC are dependent on the assumption that oligomeric species do not interconvert on their experimental timescales. Another technique, fluorescence correlation spectroscopy, can detect oligomers but not differentiate between them [99]. MS enables the real time observation of which oligomers are present in the lag phase, and how these oligomer populations change over time. Thus MS can elucidate both the self assembly processes and off pathway events from their earliest steps, although it is not possible to distinguish between these. Individual species can be observed, however transiently and low populated, in contrast to other techniques which average data over the whole population present. A recent study [100] suggests that MS can be used to quantify the heterogeneous oligomer populations present in an aggregating system, and agrees very favourably with SEC quantification data. Practical considerations such as small sample volumes, speed of data acquisition and ability to sample directly from an aggregating solution also favour MS use. Thus MS is an exciting high resolution

technique to observe and quantify oligomers. Ion mobility MS can add to the information by identifying conformational families within oligomeric orders and different oligomeric orders which appear at a coincident mass to charge ratio. The collision cross sections of oligomers can also be measured, which, combined with molecular dynamics simulations, is a powerful method for analysing conformational dynamics and structural characterisation.

1.4. Biological Mass Spectrometry

1.4.1. Introduction

The basic principle underlying all mass spectrometers is the production and injection of ions from a source which pass through a mass analyser to a detector where they are recorded according to their intensity and distinguished by their differing mass to charge ratios (*m/z*).

1.4.2. Ionisation methods

1.4.2.1. Electrospray Ionisation

The first studies of electrospray were undertaken by Zeleny in 1917 [101], and, over fifty years later, Dole used electrospray ionisation (ESI) to determine molecular weight [102]. Fenn developed this method, showing that biological molecules could be transferred into the gas-phase as multiply charged ions and their molecular weight determined [103,104], work for which he shared the Nobel Prize in 2002 [105]. ESI is now an invaluable tool for protein analysis as biological molecules can be conveyed intact from solution to vacuum [106] and non-covalent interactions preserved, allowing for the transfer of entire protein complexes into the gas-phase [107,108,109].

For the process of ESI to occur the compound being analysed must be in solution. The analyte is infused from a syringe at atmospheric pressure, nebulised, and enters the mass spectrometer as charged droplets through the entrance aperture – the 'sample cone' (Figure 1.3). Desolvation occurs from the exit of the capillary to the first vacuum pumping stage of the mass spectrometer. As with all mass spectrometers, the analyser region is kept at very low pressure. Ion transition between these two pressures is achieved using numerous differential stages of high-capacity vacuum pumps and focusing multipole lenses. A key advantage of ESI is that from droplet formation to analyte entry into the analyser, very little fragmentation of molecules occurs, a property particularly crucial in protein studies.

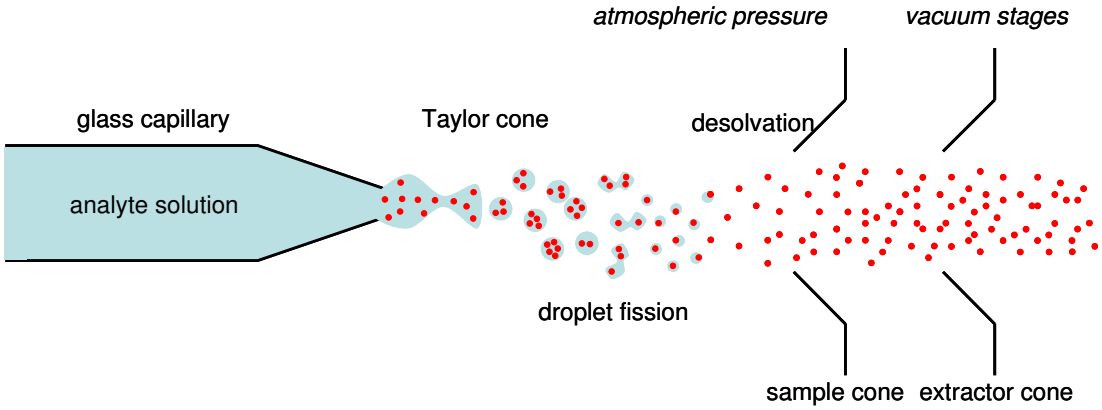


Figure 1.3 | Schematic of a n-ESI source, showing Taylor cone formation, production of multiply charged droplets, droplet fission, desolvation and entry of analyte ions into the source of a mass spectrometer through the sample cone.

An electric field is produced by applying a potential difference between the capillary and the sample cone (Figure 1.3). Typical field strengths are of the order of 10^6 Vm^{-1}, resulting in accumulation of charges (the polarity of the charges depends whether positive or negative ionisation mode is in use) at the liquid surface of the capillary. At low voltages the surface of the drop at the capillary tip appears spherical and elongates as accumulated charges in the tip exert pressure in the stronger electric field [110]. This culminates in the surface tension of the drop breaking, leading to the formation of a 'Taylor Cone', from which a plume of nebulised ions is produced

(Figure 1.3). Gomez and Tang [111] found that as droplets are deformed mechanically, the dispersion of droplets can occur before the limit imposed by the Rayleigh equation is enforced – thus reducing the necessary repulsive forces. The solvent containing the droplets now begins to evaporate, causing an increase in charge density, causing the droplets to divide further due to the interplay between cohesive and repulsive Coulombic forces. Eventually only very small droplets are left which are dispersed by a low flow rate of nitrogen gas, injected coaxially. Remaining solvent molecules are removed by passing the droplets through a curtain of heated inert gas and the heated sample cone. The desolvated analyte ion passes into the mass spectrometer through the cone, guided by decreasing voltages and pressure.

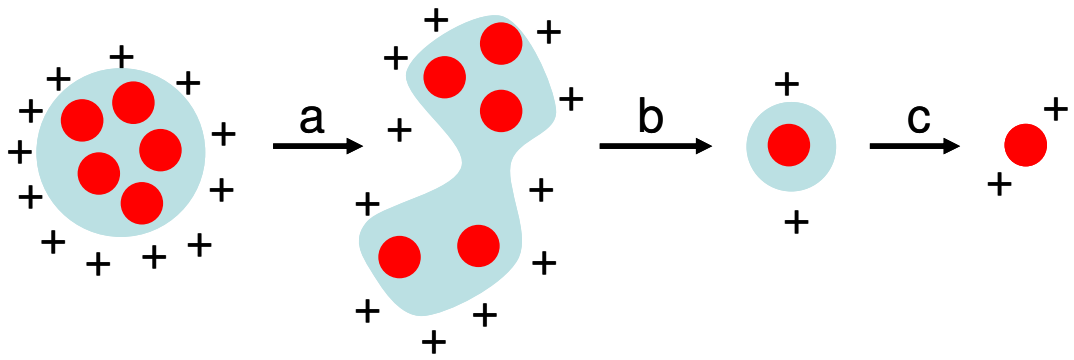


Figure 1.4 | Schematic of the charged residue model where a multiply charged and solvated cluster of ions produces single, desolvated analyte ions. ***a*** *fission of ions.* ***b*** *further fission.* ***c*** *complete desolvation.*

The process by which a charged droplet becomes a desolvated ion is believed to occur by one of two processes. The more commonly occurring process [112] is the charged residue model [102] whereby droplets undergo sequential droplet fission until single solvated analyte ions remain which are then desolvated (Figure 1.4). This model is more relevant for large biomolecules such as those which are the topic of this thesis [113]. However a competing mechanism, favoured by small ions, is the ion evaporation model [114] where analyte ions are expelled from droplets due to the presence of Coulombic repulsive forces which increase as the droplet decreases in size.

In large molecules, such as proteins, the charge observed is related to the availability of ionisable sites. This links with an important property of ESI; it produces multiply charged ions from large molecules due to charge accumulation in the droplets. Advantages of this are that it allows the analysis of molecules with a high molecular weight [106] and improves detector sensitivity. ESI spectra of proteins display consecutive peaks corresponding to multiply charged molecular ions which have been protonated or deprotonated. It is necessary to determine the molecule's original mass before assigning the peaks to particular oligomeric and charge states.

1.4.2.2. Nano-electrospray Ionisation

A technological advancement in ESI which uses much lower flow rates, sample volumes and concentrations is nano-electrospray ionisation (n-ESI) [115,116]. n-ESI utilises a glass capillary with a micron diameter tip, rather than the metal needles employed in ESI which have tip opening of circa 0.1 mm (Figure 1.3). Electric charge is supplied to n-ESI needles *via* a wire inserted into the analyte, or the capillary itself is covered with a conductive coating. Droplets produced by n-ESI are significantly smaller than those produced by ESI [116,117] which results in a less non specific aggregation and a gentler desolvation process as the number and energy of collisions necessary to desolvate the analyte molecules are fewer [118]. As droplets are originally smaller, the final droplets will contain lower concentrations of unwanted salts or impurities, resulting in a cleaner mass spectrum [118].

1.4.3. Mass Analysers

Once produced, ions are separated according to their *m/z* ratio by mass analysers. There are a number of mass analysers available, which use different principles to obtain *m/z* separation. Utilised in the experiments which follow and discussed in detail below are the quadrupole, time of flight and Fourier transform ion cyclotron

resonance (FT-ICR) analysers. Quadrupole analysers separate ions based on the differing trajectory stabilities of ions of different *m/z* values, time of flight analysers, as their name suggests, separate according to flight time and FT-ICR analysers exploit differing *m/z* resonance frequencies.

1.4.3.1. Quadrupole Analyser

A quadrupole mass analyser consists of four parallel rods to which a direct current (DC) and an alternating radio frequency (RF) potential are applied. Opposing rods have an in-phase RF potential (e.g. positive polarity) whilst adjacent rods are out of phase with respect to RF potential (e.g. negative polarity). This creates an oscillating electric field, described by Equation 1.1: Φ_0 is the potential applied to the quadrupole rods, U is the DC potential, V the zero-peak amplitude of the RF potential, ω the angular frequency of the RF potential and t is time.

$$\pm\Phi_0 = \pm(U - V\cos\omega t) \qquad \textit{Equation 1.1}$$

Ions entering the quadrupole will be attracted to a rod of opposite polarity. However as the polarity of the rods is continually changing due to the alternating RF potential ions prescribe an oscillating path. Ion trajectory through the quadrupole is therefore dependent on the values of U, V and ω. Ions of a specific *m/z* have specific values of U, V and ω which will enable a stable path through the quadrupole (resonant ion), and if these variables are fixed at these values all other ions will possess unstable trajectories in the oscillating field causing them to collide with the quadrupole rods and not reach the detector (non resonant ion) [119]. In this way the quadrupole can be used to select ions of a specific *m/z* to transmit, a function utilised when performing collision induced dissociation (CID). Therefore V is scanned through a range of amplitudes to allow ions with multiple *m/z* values to pass through the quadrupole. Figure 1.5 displays a typical quadrupole analyser.

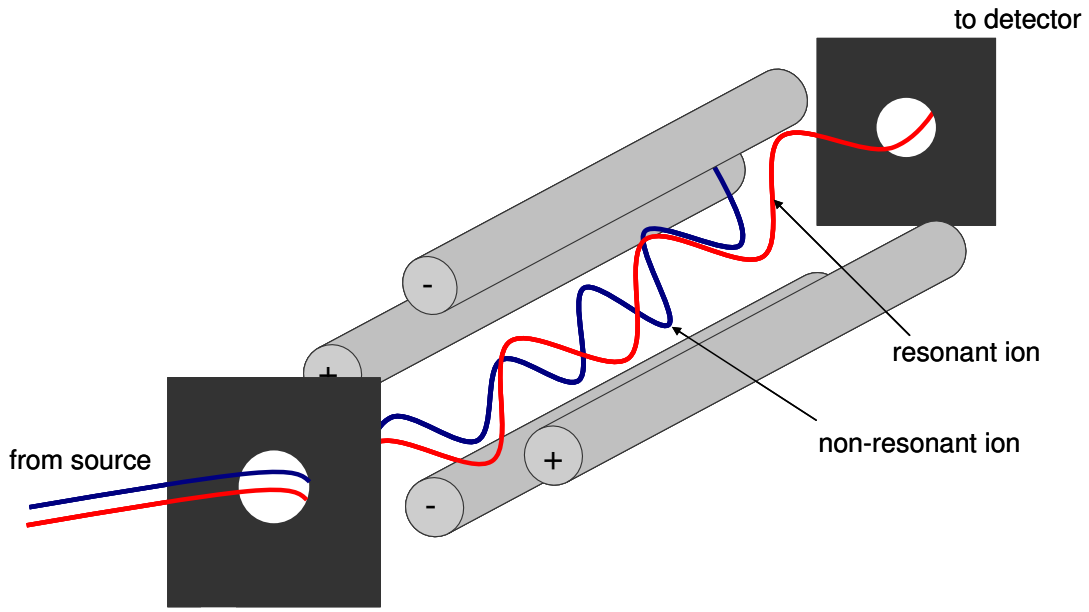


Figure 1.5 | Schematic of a quadrupole analyser, consisting of four perfectly parallel (shown) or hyperbolic rods with a circular cross section. Adjacent rods have the opposite sign. Resonant ions (red) have a stable trajectory through the analyser; non-resonant ions (blue) collide with the quadrupole rods and discharge.

Quadrupoles operating in only RF mode can focus an ion beam by creating a potential well along their central axis. This has led to the development of hexapoles and octapoles, which are utilised as ion guides in many modern mass spectrometers to transfer ions between different sections and vacuum gradients. Identical principles apply to those described above, although the shape of the potential well differs according to the number of rods present. Quadrupoles have the greatest focussing power and transmission efficiency, but over a narrower mass range than hexapoles and octapoles [110].

1.4.3.2. Time of Flight Analyser

Conceptually described in 1946 [120] and commercially developed in 1955 [121], time of flight (TOF) mass analysers accelerate ion packets into a drift tube *via* a potential difference, or pusher pulse. Following this they enter a field-free region in which their velocity is dependent on their *m/z* ratio. Therefore their *m/z* can be inferred from the time taken for them to move between the pusher pulse and the detector (Equation 1.3) [120,121].

Leaving the source, the kinetic energy (E_k) of an ion with mass (m) and total charge (ze), accelerated through a voltage (V), is given by:

$$\frac{mv^2}{2} = zeV = E_k \qquad \textit{Equation 1.1}$$

From basic equations of motion the time taken (t_f) for the ion to travel the distance d at velocity v is

$$t_f = \frac{d}{v} \qquad \textit{Equation 1.2}$$

Thus

$$t_f{}^2 = \frac{m}{z}\left(\frac{d^2}{2eV}\right) \qquad \textit{Equation 1.3}$$

Equation 1.3 demonstrates how the *m/z* ratio of the ion can be calculated from its time of flight. There is no theoretical upper limit to the mass range which can be analysed by a time of flight mass analyser.

However the closer the ion's proximity to the accelerating voltage the greater the potential difference it experiences. Therefore ions of the same *m/z* ratio could enter the drift tube with differing kinetic energies, contrary to the equations above. The resolution of original linear drift tubes suffered as a result of this as these ions would

arrive at the detector at slightly different times. This problem was solved by the use of a reflectron (Figure 1.6).

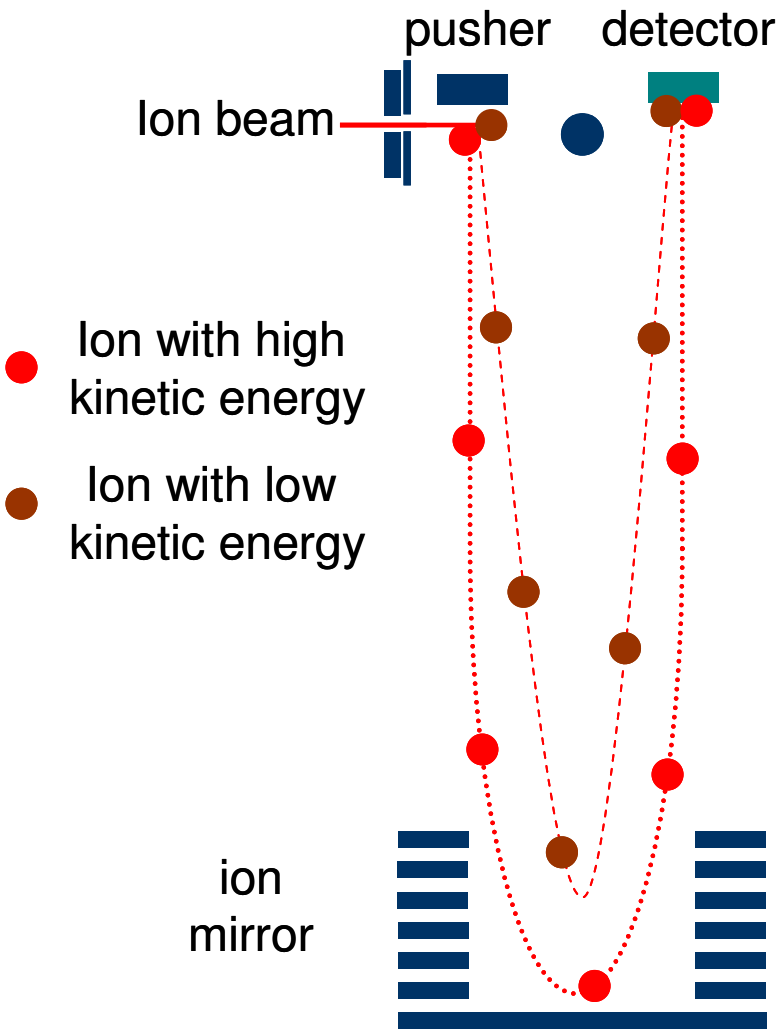


Figure 1.6 | Schematic of a TOF mass analyser. Ions with identical mass but differing kinetic energy entering the TOF arrive at the detector at the same time. This improves mass resolution by eliminating differences in the energy of ions of the same m/z ratio.

A reflectron [122] is a series of ring electrodes which create a retarding field, acting as an 'ion mirror' to reflect ions back towards the detector. Ions which leave the pusher with a higher kinetic energy travel further into the reflectron than those with a lower

kinetic energy but identical mass. Thus the distance ions with a high kinetic energy travel inside the reflectron is further; adjusting the ion mirror enables ions of the same *m/z* ratio to arrive at the detector at the same time, regardless of their kinetic energy. This, coupled with the effective doubling of the length of the TOF tube, increases instrument resolution dramatically, although with the penalty of introducing a mass limitation.

1.4.3.3. FT-ICR Analyser

Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS was developed in 1974 [123] and a comprehensive review has been published subsequently [124]. FT-ICR MS can obtain very accurate mass measurements, and due to its high resolution, isotopic baseline resolution of high charge states. It is based on the principle that ions in a uniform magnetic field will experience a force perpendicular to the direction of the applied magnetic field. This is known as the Lorentz force and is described by Equation 1.4: *m* is mass, *a* is acceleration, *z* is charge, *v* is velocity and *B* is magnetic field

$$\vec{F} = m\vec{a} = m\frac{d\vec{v}}{dt} = z\vec{v} \times \vec{B} \qquad \textit{Equation 1.4}$$

Ions travelling at a constant speed will be diverted by the Lorentz force into an arc - cyclotron motion. As the Lorentz force acts as a centripetal force and can be described by Equation 1.5 where *m* is mass, *r* is radius of ion path curvature, *z* is charge, *v* is velocity and *B* is magnetic field.

$$\frac{mv^2}{r} = zvB \qquad \textit{Equation 1.5}$$

$$\omega = \frac{v}{r} \qquad \textit{Equation 1.6}$$

Further manipulation of Equation 1.5 and Equation 1.6, where ω is angular velocity, results in:

$$m\omega r^2 = zB\omega r \qquad \textit{Equation 1.7}$$

This can be transformed into:

$$\omega = \frac{zB}{m} \qquad \textit{Equation 1.8}$$

Thus ions of the same *m/z* ratio, regardless of their kinetic energy, have the same ICR frequency. In practice, ions are pulsed into the ICR cell where they are trapped by confining potentials applied to the trapping plates. Ions can be stored in the ICR cell for hours, where they undergo motion due to the magnetic and electrical forces they experience. This motion is best described as cyclotron motion around the centre of the ICR cell.

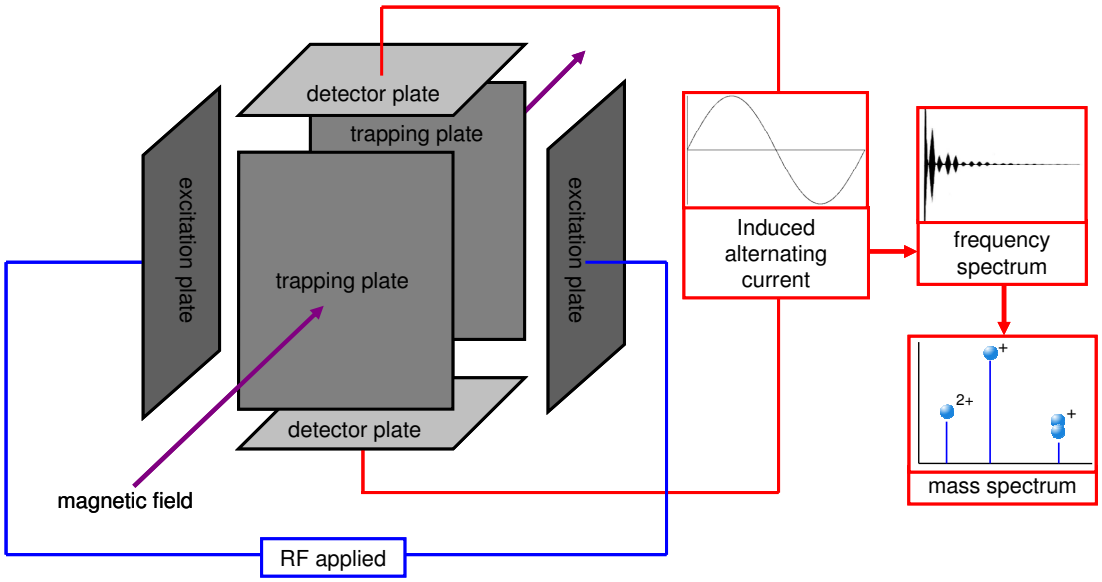


Figure 1.7 | Schematic of an FT-ICR cell. The ion undergoes cyclotron motion orthogonal to the magnetic field.

Ions undergoing cyclotron motion do not produce any detectable signal. The ions must be excited using a resonant frequency, applied to parallel plates of the ICR cell (which are perpendicular to the magnetic field direction), to obtain larger cyclotron

radius [125]. Ions of same *m/z* move in coherent packets; these packets, passing near the electrodes, induce a charge in that electrode. This creates an alternating current in the detector circuit and this signal is amplified and digitised. An excitation frequency sweep (broadband excitation) allows ions of multiple *m/z* values to increase their orbit and be detected [123]. This is in contrast to other mass spectrometry instrumentation where ions are destroyed in the process signal detection. Resolution is limited by length of acquisition time (as all ions are detected simultaneously) and the strength of the magnetic field [124].

1.4.4. Detectors

Mass spectrometry detectors are devices which detect ions and create an electric current of corresponding amplitude. There are several types of commonly used detectors, a comprehensive review of which can be found in 'Metastable Ions' by Cooks [126] but is beyond the scope of this thesis. The detectors utilised in the experiments in the following chapters are described below.

1.4.4.1. Photomultiplier Detector

If only the quadrupole mass analyser is used, ions are detected on their entry to the TOF tube by a photomultiplier (Figure 1.8). The incident ion beam strikes a conversion dynode which emits electrons. These electrons then collide with a phosphor screen which emits photons. These photons pass into a photomultiplier tube, where they produce an electric current which can be amplified [110].

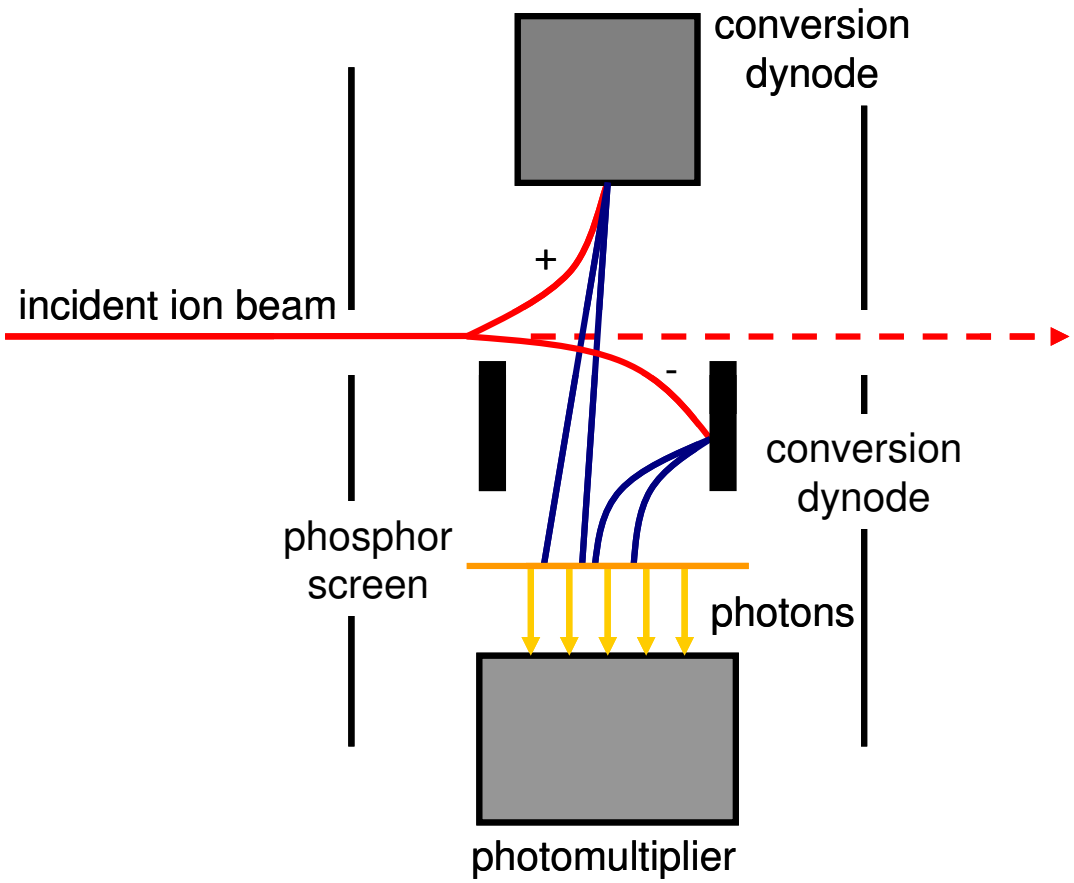


Figure 1.8 | Schematic of the photomultiplier detector. Red dotted line indicates the path of the ion when the detector is not used.

1.4.4.2. Microchannel Plate Detector

Microchannel plate detectors (MCPs) consist of many miniature electron multiplier channels (Figure 1.9). When an ion collides with the semiconductor surface on a microchannel, electrons are ejected. These electrons also strike the microchannel walls, displacing more electrons. This process is repeated causing a cascade of electrons travelling towards the ground potential which are accelerated into a series of continuous dynode electron multipliers, as each microchannel can be thought of as a continuous dynode structure [127]. This progression results in a measurable current

corresponding to the detection of one ion. An amplifying effect of up to 10^8 can be achieved if several MCP plates are used [110].

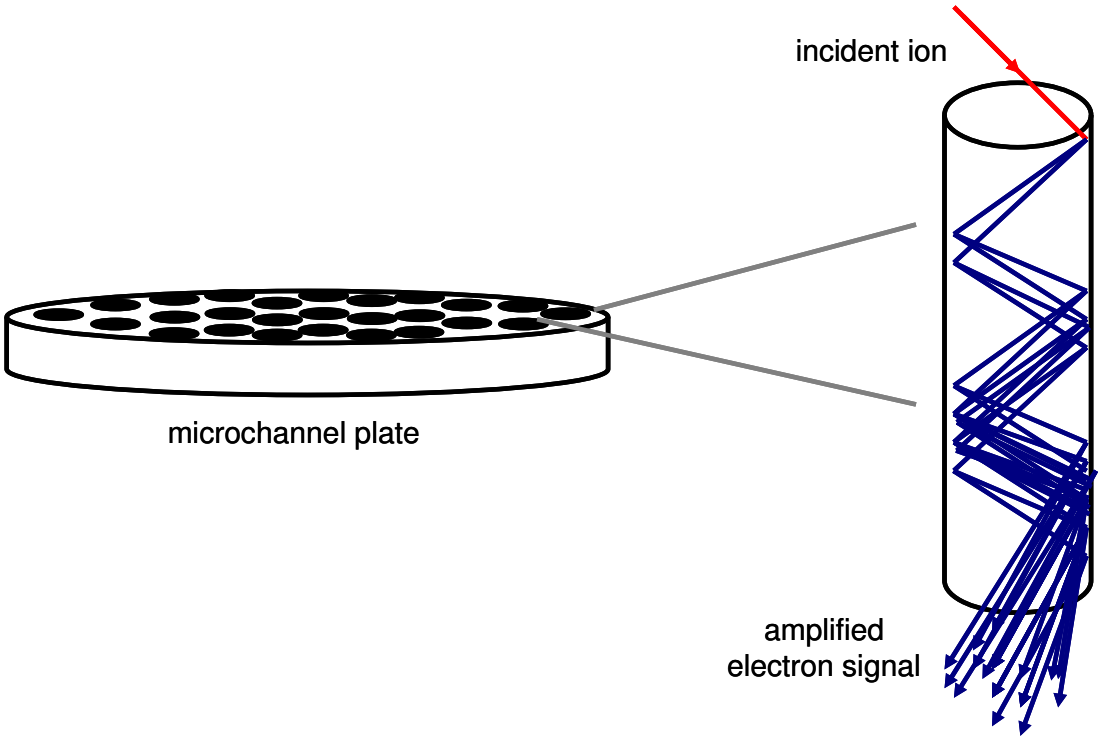


Figure 1.9 | Schematic of a MCP detector, with an enlargement of one microchannel.

Short secondary electron path lengths in the channels enable a quick response, allowing the determination of precise arrival times and correspondingly high signal resolution.

1.4.4.3. FT-ICR MS Detector

A broadband frequency sweep produces a signal in the detector plates which is a superposition of sine waves of different frequencies and amplitudes induced by the proximity of different *m/z* ion packets. Applying a Fourier transform to this results in a spectrum of the component cyclotron frequencies and their respective amplitudes. Cyclotron frequency can then be converted into a conventional *m/z* versus ion intensity spectrum using Equation 1.8.

1.5. Ion Mobility Mass Spectrometry

1.5.1. Introduction

Ion Mobility Mass Spectrometry (IM-MS) is a gas-phase electrophoretic separation technique which differentiates between ions based on their mass, charge and mobility. Historical investigations into the movement of ions in different gases [128,129] led to the development of the first drift tube where ions travel through a buffer gas filled tube under the influence of a weak electric field [130]. Combining ion mobility spectrometry (IMS) with MS instrumentation created a powerful new analytical technique [131]. Since its conception in 1963 the field of IM-MS has blossomed, with a significant growth in publications utilising this technique [132,133]. There are three types of IM-MS instrumentation most commonly used in industry and research groups; Field Asymmetric Ion Mobility Spectrometry (FAIMS), Travelling Wave Ion Guides (TWIGS) and Drift Tube Ion Mobility Mass Spectrometry (DT-IM-MS). Of these DT-IM-MS most closely resembles historical IMS apparatus. All IM-MS experiments performed in this thesis utilised DT-IM-MS, which is described in detail below. FAIMS and TWIGS are briefly described, comprehensive reviews of all IM-MS instrumentation can be found in both Hill *et al.* [134] and Harvey *et al.* [135].

1.5.1.1. DT-IM-MS

By measuring the time taken for ions to move through a drift cell, filled with a buffer gas, under the influence of a weak electric field, it is possible to determine the rotationally averaged collision cross sections of the ions present. The larger the cross section of the ion available for collisions with the gas atoms (known as the collision cross section or CCS), the slower that ion will drift through the cell. Small, compact ions will travel more quickly than larger ions with an extended conformation (Figure 1.10). Ion intensity is recorded as a function of time, in a plot called an arrival time

distribution (ATD). This measures the mobility of the ion which provides details concerning structure.

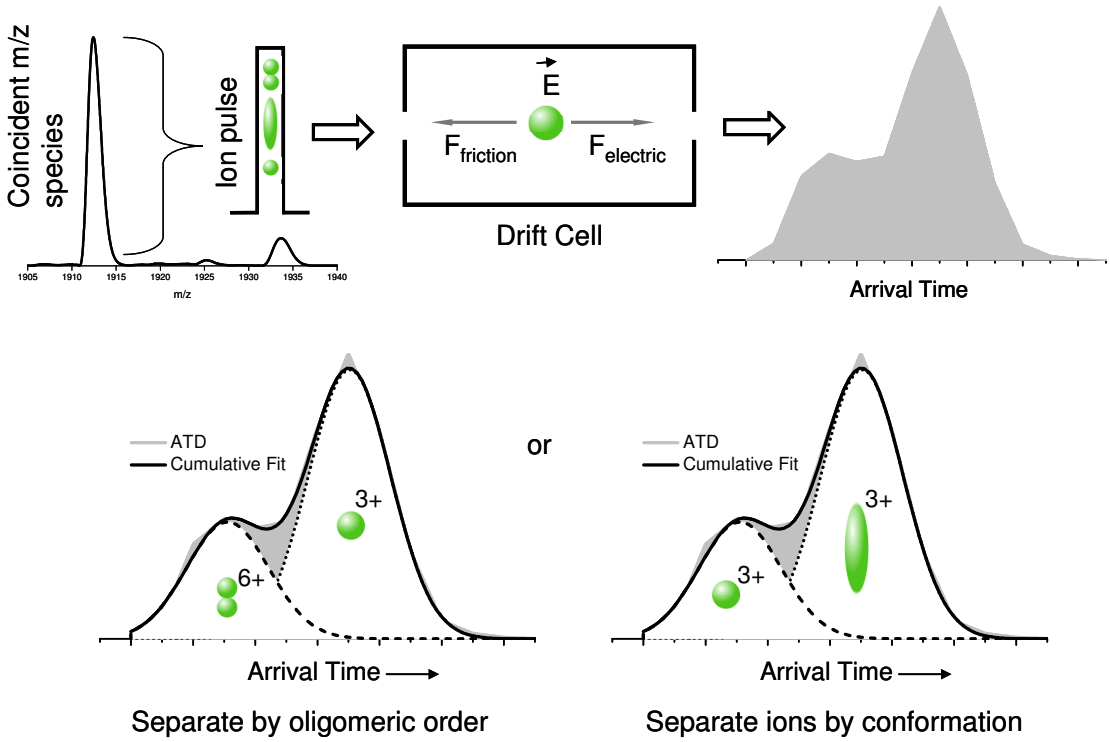


Figure 1.10 | IM-MS separation: a diagram showing how the charge and conformation of ions influences their arrival time distribution.

The electric field can be described in terms of the length of the drift cell and the voltage applied across the cell by Equation 1.9: E is the electric field across the cell, L is the length of the drift cell and V is the potential applied across the cell.

$$E = \frac{V}{L} \qquad \textit{Equation 1.9}$$

The ratio of the electric field applied across the cell to the buffer gas number density (E/N) determines the motion of an ion through the cell. The physical unit of the ratio E/N is called a Townsend (Td) and is defined as 1 $Td = 10^{-17}$ Vcm^2. At high values of E/N ions align in the electric field and their mobility is dependent on E. Values of E/N which are ≤ 10 Td are known as the low field limit; in the low field limit ion

velocity is constant and linearly proportional to the electric field [136], as described by Equation 1.10 where v_d is drift velocity and E is electric field. The constant of proportionality, K, is the low field ion mobility.

$$v_d = K \cdot E \qquad \textit{Equation 1.10}$$

Low field ion mobility is proportional to buffer gas pressure, inversely proportional to ion CCS and can be converted to the 'reduced' mobility K_0 by Equation 1.11 where normalised temperature T_0 and pressure P_0 are 273.15 K and 760 Torr respectively.

$$K_0 = K \frac{P}{P_o} \frac{T_0}{T} \qquad \textit{Equation 1.11}$$

The equation for the time spent by the ions in the drift cell, t_d, can therefore be calculated:

$$t_d = \frac{L}{v_d} = \frac{L}{K \cdot E} = \frac{P}{V} \frac{L^2 T_0}{K_0 P_0 T} \qquad \textit{Equation 1.12}$$

Arrival time, t_a, measured by the ATD, is the time from ion injection into the cell until detection. It is equal to the time taken for the ions to drift through the cell, t_d, plus the time the ions spent outside the drift cell (known as the dead time t_0).

$$t_a = t_o + t_d \qquad \textit{Equation 1.13}$$

If the temperature of the drift cell is kept constant, a graph of arrival time against P/V will give a linear relationship with an intercept equal to the dead time. The gradient of the line of best fit is equal to the inverse of the reduced ion mobility, as the values of L and T in the equation are known.

The relationship between reduced ion mobility and the ion's CCS is described by Equation 1.14: z is ion charge, e is electron charge, N_0 is the buffer gas number

density, μ is the reduced mass of the buffer gas and ion, k_b is the Boltzmann constant, T is the effective temperature and Ω is the momentum transfer collision integral.

$$K_0 = \frac{3ze}{16N_0}\left(\frac{2\pi}{\mu k_b T}\right)^{\frac{1}{2}}\frac{1}{\Omega} \qquad \textit{Equation 1.14}$$

By measuring the ATD and obtaining the reduced ion mobility, the momentum transfer collision integral, Ω, can be calculated. Ω, by kinetic theory, can be approximated to the CCS of the ion σ. For most geometric shapes of ion, the differences between Ω and σ can be corrected for empirically. CCSs are dependent on the buffer gas present in the drift cell [137,138]. The resolution of DT-IM-MS depends on gas pressure and the length of the drift cell, as increasing the time ions spend in the drift cell allows the greatest mobility separation. With DT-IM-MS it is possible to achieve the highest resolution of any IM-MS technique [139].

1.5.1.2. TWIGS

Introduction of commercially available TWIGS instruments; the Synapt G1 [140], G2, and as of 2012, the G2-S, has expanded the number of research groups with the capability to use IM-MS. In brief TWIGS instruments use a series of stacked ring RF ion guides [141], combined with a travelling wave of voltage, to separate ions. Ions are stored in the first ion guide before injection into a mobility separator where a travelling wave of transient DC voltages are applied in conjunction with an RF voltage. Ions with highest mobility 'ride' the crest of the travelling wave, whilst lower mobility ions fall behind successive wave fronts before reaching the end of the separator. Upon exiting the separator the third ion guide conveys the ions to the TOF compartment. There is no theoretical equation relating the mobility of ions in TWIGS devices and their CCSs [142], therefore a calibration must be applied using standards whose CCSs, determined by DT-IM-MS, are known [143,144]. However parameters on commercial TWIGS instruments must be carefully controlled to prevent excessive heating of ions, which can induce conformational changes [145].

Thus issues of calibration and ion heating remain a caveat in comparing TWIGS gas-phase structures and CCSs to solution phase structures.

1.5.1.3. FAIMS

FAIMS, in opposition to traditional DT-IM-MS, utilises high field mobility of ions at atmospheric pressure [146,147]. Ions travel between two electrodes against a flow of gas; one electrode is grounded whilst an asymmetric waveform constituting long low voltages and short, opposite polarity, high voltages is applied to the second. The electric field the ions experience is orthogonal to the gas flow and causes them to follow oscillating paths. To prevent collision with the electrodes a compensating DC voltage is applied. The ratio of an ion's high field to low field mobility determines which value of the compensation voltage will allow the ion to pass through the device [148], hence allowing the separation of different ions. An ion's low field mobility, and hence its CCS, cannot be determined from the FAIMS technique; however FAIMS is a powerful separation technique alone, and when combined with DT-IM-MS can provide an additional separation dimension. Commercial instruments are available from Thermo Scientific, Ionalytics and Agilent Technologies in collaboration with Owlstone.

1.6. Molecular Modelling

Theoretical CCSs can be calculated using molecular coordinates obtained from an array of biophysical techniques such as NMR, X-ray crystallography and molecular dynamics simulations. The comparison of experimentally derived CCSs with these theoretical CCSs can allow structural assignments, or at least, indications of structure to be ascertained.

MOBCAL [149,150,151] can calculate CCS from molecular coordinates using one of three methods; exact hard sphere scattering (EHSS), projection approximation (PA) or the

trajectory method (TM). The PA is the least computationally demanding as each molecular atom is modelled as a hard sphere with interactions between buffer gas and ions also as hard sphere collisions; the resulting CCS is averaged over all possible orientations of the ion. An improvement on the PA is the EHSS model which includes terms for multiple collisions and scattering. The most representative model of the interactions between buffer gas atoms and ions is the TM, which considers the analyte ion as a collection of molecules, possible multiple collisions and the long range interactions which occur between drift gas atoms and the ion. However, as a result of this it is the most computationally expensive. For small convex ions the PA provides a sufficiently accurate representation of CCS, however for proteins it drastically underestimates the CCS. For these more complex structures, EHSS and TM are more accurate estimators of CCS [152].

1.7. Use of MS and IM-MS to Study Biological Molecules

As MS and IM-MS measurements of biological molecules are performed in the gas-phase the relevance of these gas-phase structures to solution phase structures is an important question. Initial evidence comes from the fact that the mass spectrum of a protein is influenced by its solution environment prior to ionisation and entry into a mass spectrometer [153,154]. Denaturing conditions produce a larger and more highly charged envelope of species than those produced from a 'native' solution, suggesting that the protein has unfolded under these solution conditions, and these unfolded species are preserved into the gas-phase. It could also be argued that the gas-phase environment of a mass spectrometer, where the dielectric constant is 1, is more similar to that proteins experience in their cellular environment (from 2 - 6 [155]) than in a dilute aqueous solution (dielectric constant 78 [156]) [157]. Upon desolvation a loss of hydrophobic interactions will occur and electrostatic interactions will be strengthened. Despite these changes, soft landing experiments on lysozyme and trypsin found that their biological activity was unimpaired by their passage through a

mass spectrometer, lending support to the relevance of MS measurements to biological structures [158,159].

Evidence for the retention of non-covalently bound solution structures in the gas-phase came in 1991 with the detection of receptor-ligand [108], enzyme-substrate [107], and haemoglobin complexes [109]. Robinson *et al.* [160] performed one of the first studies to provide direct evidence that protein quaternary complexes could remain in their native form in a mass spectrometer. Using a TWIG instrument, CCSs of assemblies of the tryptophan RNA binding protein, TRAP, were determined. X-ray crystallography has shown that TRAP forms a ring topology of 11 subunits [161]. MS of the unbound apo form of TRAP revealed a charge state envelope of 22+ to 19+ ions. IM-MS measurements revealed that the conformation of the assembly is dependent on charge state, although the charge is below the limit calculated to cause coulombic unfolding. The 19+ ion presents predominantly in one conformation, the CCS of which agrees with the calculated CCS of the X-ray ring structure. As these experiments were performed in the early days of TWIGS, and corresponding early calibration protocols, absolute CCS values should be treated with caution. CCSs of the 20+ ion exhibit a bimodal distribution, indicating considerable structural heterogeneity; from a ring topology to more collapsed conformations. The two most highly charged ions, 21+ and 22+, have compact CCS distributions which, in conjunction with modelling results, are postulated to be very compact structures with closely packed subunits. Thus the 11-membered ring topology of the complex was postulated to have been preserved, *in vacuo* and without the presence of bulk water.

Stability of the TRAP ring complex is known to increase on binding to tryptophan (Trp) [161], although the MS spectra of apo TRAP and TRAP in the presence of Trp appear very similar. The maximum CCS experimentally observed for both apo and Trp bound form are similar, from which it can be inferred that Trp bound TRAP retains its ring topology. However, for the 20+ ion the conformational heterogeneity of the apo TRAP ion is converted into a single conformation which is of a CCS consistent with the ring structure. From this difference it can be inferred that the

postulated ring structure of TRAP is indeed more stable in the presence of Trp even in the absence of bulk solvent.

Many aspects of protein structure can be preserved in the absence of solvent, including secondary, quaternary and/or compact structure and hydrogen bonding interactions [162]. It is not expected that identical interactions will be present in the presence and absence of bulk solvent, but that the overall size and majority of the binding interactions found in solution can be successfully preserved *in vacuo*. Breuker and McLafferty [163] evaluate the temporal evolution of protein structure which occurs on transfer into the gas-phase from experimental evidence and molecular dynamics (MD) simulations. On the 10^{-12} seconds timescale after desolvation, charged side chains collapse which can transiently stabilise the native fold of a large, compact proteins, whilst on the 10^{-3} seconds timescale, loss of hydrophobic interactions and dissociation of electrostatic bonds can occur. New non-covalent bonds can be formed within seconds to minutes. However ionic interactions and those of residues buried deep in the core of the protein are unlikely to be affected by desolvation.

1.7.1. Amyloidogenic Peptides and Proteins Studied by MS, IM-MS and Modelling

A growing body of work assessing the configurations of species formed during aggregation has utilised MS [13] and IM-MS techniques, often combined with simulations; a brief review follows of those systems which have been investigated most extensively.

1.7.1.1. Islet Amyloid Polypeptide

The aggregation of human islet amyloid polypeptide (hIAPP) into amyloid fibrils is implicated in the pathogenesis of type II diabetes [164]. However, the rat homologue,

rIAPP, which differs by six residues in the 'core mutation region', does not form amyloid fibrils [165]. The sequences of hIAPP and rIAPP are shown below with the differing residues highlighted.

hIAPP: KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY

rIAPP: KCNTATCATQRLANFLVRSSNNLGPVLPPTNVGSNTY

This provides an opportunity to compare the non-amyloidogenic oligomeric species formed by rIAPP and the on or off pathway oligomers present in the lag phase of hIAPP fibrillation [27]. At pH 6.4 mass spectra of rIAPP and hIAPP have almost identical charge state distributions; a dominant $[M+3H]^{3+}$ peak, a less intense $[M+4H]^{4+}$ and a small signal from a $[2M+5H]^{5+}$ species (Figure 1.11A) [166]. However important conformational differences can be divulged from the monomer ATDs. For the dominant 4+ charge state hIAPP contained a later arriving ATD peak in addition to the two earlier arriving peaks which are shared with rIAPP. Further studies implied that this extended conformer originated from a solution structure family rather than being an experimental artefact. Replica exchange MD (REMD) simulations of rIAPP and hIAPP monomers identify the mutual compact structures as helix-coil, whilst the unique hIAPP extended conformation is suggested to correspond to a β-hairpin structural family. Secondary structure sequence analysis confirmed that sequence variations between rIAPP and hIAPP have a drastic effect on peptide conformation, with the 'core mutation region' adopting coiled and β-strand structures respectively.

The population of this β-hairpin family by hIAPP could be responsible for its amyloidogenic nature, supported by the absence of any β-hairpin conformer in rIAPP experiments and simulations [166], and similarities with a solid-state NMR structure [167]. Study of the 5+ dimeric species which is present in both rIAPP and hIAPP at lower intensities than the monomeric species (Figure 1.11A) [168] revealed multiple conformational families (Figure 1.11B). The most populated hIAPP dimer conformation is 12% larger than the rIAPP dimer, consistent with the hIAPP β-hairpin monomers which were 17% larger than the coil-rich rIAPP monomers [166]. It is therefore suggested that hIAPP dimers could be composed of β-hairpin monomers.

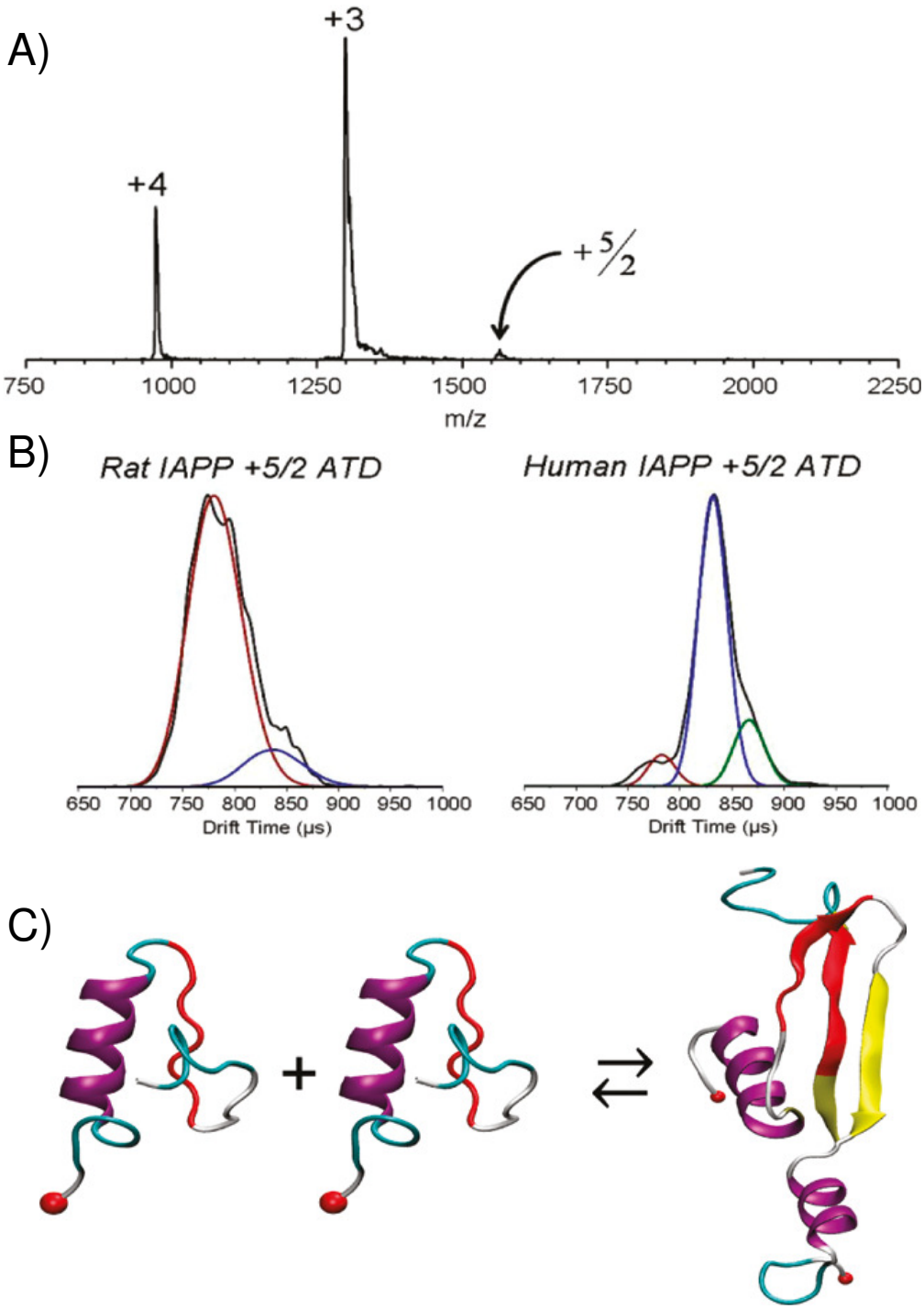


Figure 1.11 | IAPP A) n-ESI mass spectrum of hIAPP. (B) Rat and Human IAPP dimer ATDs contain multiple features. C) hIAPP trajectory shows conversion of coil into β-strand upon dimerisation. Figure and caption from Dupuis et al. [168].

MD simulations of dimers suggest that rIAPP forms disordered compact dimers from coil-rich monomers, whereas hIAPP dimers are more extended and β-strand rich. The interaction of hIAPP monomers occurs between β-strand conformational areas containing the mutation region and side by side assembly of β-hairpins was the dominant binding mode [168]. Simulations have also shown some evidence that hIAPP β-strand rich monomer interfaces can recruit coil and helix regions in during the dimerisation process. Thus hIAPP extended β-hairpin monomers form extended β-strand rich dimers (Figure 1.11C), in contrast to the rIAPP formation of compact, coil rich dimers [166]. To summarise, IM-MS and MD have elucidated a possible mechanism for the ability of hIAPP to aggregate, contrasted with unamyloidogenic rIAPP.

1.7.1.2. β_2-microglobulin

β_2-microglobulin (β_2m), is part of the major histocompatibility complex class 1 molecule [169,170], which protects native cells from attacks by T lymphocytes [171]. A complication of long term haemodialysis is dialysis-related amyloidosis, where fibrils largely composed of β_2m form in tissues. This is believed to be due to the high concentrations of circulating β_2m which build up in the bloodstream as they are not removed by haemodialysis [172].

The presence of different conformers of a protein can be inferred from the existence of different charge state distributions in an ESI-MS spectrum [173]. This allowed Ashcroft *et al.* to follow the pH driven unfolding of β_2m *via* a broadening of charge state distribution and a shift to more highly charged protein ions and compare the behaviour of wild type (WT) β_2m with that of mutants [174]. Three conformational families of β_2m monomer were identified at pH < 5; a native, partially folded and an acid unfolded state [174,175], and their presence confirmed by CD and NMR [176].

IM-MS has significantly advanced research into the early aggregation *in vitro* of β_2m. It allows the unambiguous separation of conformers prior to mass analysis and

thus makes superfluous the deconvolution of charge state distributions. The conformational families of oligomers can be quantified in terms of shape, dynamics and population [177,178] and monitored in different solution conditions.

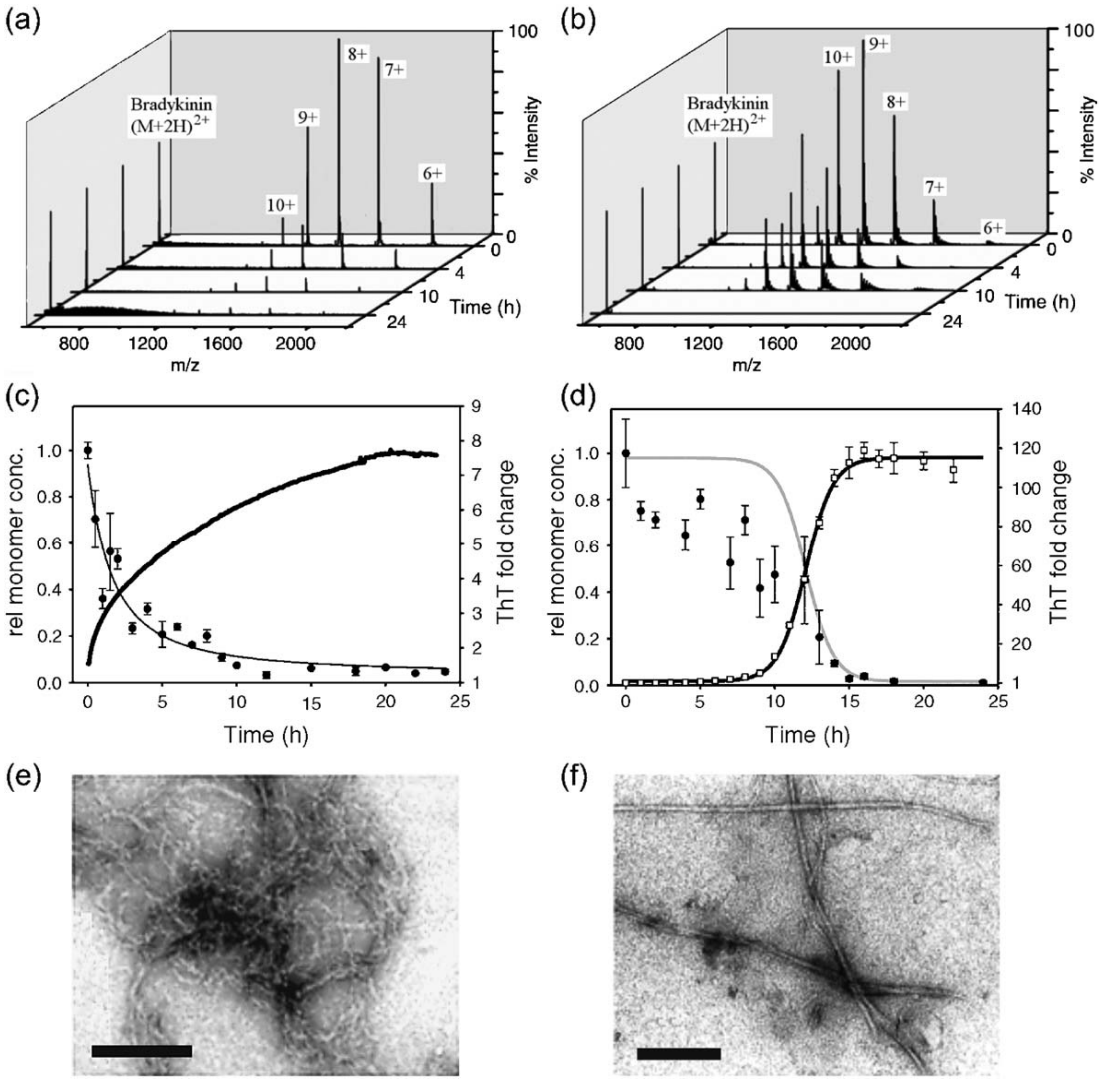


Figure 1.12 | β_2m (a) and (b) Time course mass spectra of aggregating 17 μM β_2m at 37 °C of in (a) 150 mM ammonium formate at pH 3.6 and (b) in 100 mM ammonium formate pH 2.5 (shaken at 200 rpm). (c) and (d) Fibril growth kinetics at (c) pH 3.6 and (d) pH 2.5 monitored using Thioflavin T fluorescence (continuous line in (c) and open squares in (d)). Concentration of monomeric β_2m measured by nESI-MS is shown in (c) and (d) as filled circles (error bars represent one standard deviation from the mean). (e) and

(f) EM images of fibrils formed at (e) pH 3.6 and (f) pH 2.5 (scale bar 200 nm). Figure and caption from Smith et al. [91]

Under different conditions *in vitro* two fibril morphologies have been observed [70,91,179]. Low pH and ionic strength solution conditions result in the formation of long, straight (LS) fibrils (Figure 1.12f) *via* a nucleation mechanism, conversely higher pH and ionic buffer strength result in the spontaneous assembly of shorter worm like (WL) curved fibrils (Figure 1.12e). During aggregation β_2m monomer depletion and corresponding evolution of oligomers can be observed by ESI-MS [91] for both fibril species (Figure 1.12a,b). In the lag phase of LS fibril assembly, at pH 2.5, the acid unfolded monomeric conformation is the most highly populated, with dimer, trimer and tetramer species also detected by ESI-IM-MS [91,180]. Calculation comparison with experimental CCSs suggest that monomers assemble in a elongated linear fashion for LS fibril formation [180].

At pH 3.6 β_2m is chiefly a partially folded monomer (as measured by NMR); additionally less abundant oligomers of dimer to hexameric order are present [91,174,179]. When the ionic buffer strength is increased to initiate WL fibril growth, oligomers of dimeric to nonomaric are detected. ^{14}N- and ^{15}N-β_2m exchange experiments show dynamic and rapid oligomeric subunit exchange. When exchange rates for trimer, tetramer and pentamer decrease, oligomers of >14-mer are observed. It was postulated that the CCS values of these suggested that, instead of a globular shape, the oligomers consist of stacked monomers held together by edge on edge interactions which retain a β-sheet native structure [180]. Under both LS and WL fibril conditions, high charge state trimers display fast subunit exchange whilst more compact, lower charge state trimers are more stable [180,181]. Tetramers present in the LS fibril assembly pathway are less compact and more highly charged than the compact tetramers which are present only in WL fibril assembly [180]. It is hypothesised that an increase in partially folded monomer populations could lead to compact tetramers, which might catalyse the spontaneous WL fibrillation. This implication from this hypothesis is that monomer conformation has an effect on mature fibril morphology [180].

1.7.1.3. α-synuclein

Aggregated α-synuclein is the predominant component in Lewy bodies and neuritis, which are characteristic of Parkinson's disease and related synucleinopathies [182], and is implicated in the pathogenic process which culminates in Parkinson's disease. Protofibril assemblies of α-synuclein cause membrane leakage and destroy vesicular membranes [183] suggesting that non-fibrillar oligomeric species are cytotoxic species for neurons. Frimpong *et al.* recently performed a MS study of α-synuclein under conditions which were unlikely to cause amyloid formation *in vitro* [184]. They observed four distinct conformational monomer states, from random coil to highly compact structures; a conformational heterogeneity which has been observed in other studies [185]. Compact structures were consistent across a pH range of 2.5 to 10 and an ethanol content of 0 to 60%, but were most populous in acidic conditions, whilst partially folded conformations were present in significant percentages of alcohol [184]. CD experiments have been utilised to analyse the different secondary structure present in α-synuclein under different experimental conditions; at low pH α-synuclein displayed increased α-helical content, whilst the presence of ethanol increased the quantity of β-sheet present [186]. These experiments led Frimpong *et al*. to postulate that the compact structures contained increased α-helical structure whilst partially folded conformations contained high β-sheet content.

Bernstein *et al.* observe pH dependent conformations of α-synuclein ranging from compact monomers at pH 2.5 to extended monomeric conformations at neutral pH [187]. This is explained by the intrinsically disordered structure of α-synuclein at neutral pH and its high net negative charge. Protonation of the acidic C-terminus at pH 2.5 decreases the net charge and causes partial folding, a supposition which is supported by solution phase studies which suggest the formation of a partially folded intermediate is the nucleating step to fibrillogenesis [188]. It is postulated by Frimpong *et al*. that the partial folding at low pH could be α-helical in nature [184]. When experimental CCS measurements of monomeric charge states $[M-6H]^{6-}$ to $[M-11H]^{11-}$ were compared to molecular modelling structures, $[M-6H]^{6-}$, $[M-7H]^{7-}$ and $[M-8H]^{8-}$ compared favourably to compact globular structures, whereas higher

charge state CCSs agreed with elongated structures. Injection energy experiments have established that these elongated structures are not generated by gas-phase unfolding events [187].

A family of compact monomers are observed in the ATDs of Parkinson's disease related mutant A53T and WT α-synuclein species $[M-6H]^{6-}$, $[M-7H]^{7-}$ and $[M-8H]^{8-}$ [189]. Conversely another mutant A30P, also related with early onset Parkinson's, displays only a compact species in the ATD of $[M-6H]^{6-}$. This suggests that the A30P mutation causes a structural enlargement, preventing the formation of a compact structure when the net charge greater than -6. Spermine is a polyamine, present in neurons, which binds to α-synuclein at the C-terminus and is known to increase the aggregation rate [75]. The addition of spermine to the WT and two mutants at pH 7 resulted in a dominant -6 spermine-α-synuclein complex peak and a reduction of the dominant free α-synuclein charge state from -10 to -6 [189]. From ATD space it is visible that the compact monomers are more highly populated in the presence of spermine. The increased aggregation rate of α-synuclein in the presence of spermine implies that these compact forms are growth competent and quickly form amyloid fibrils.

1.7.1.4. Prions

Prions are glycoproteins which can undergo a conformational transformation from the normal α-helical cellular form PrP^{C} to a scrapie form PrP^{Sc} [190]. PrP^{Sc} has a high aggregation propensity, enriched β-sheet structure, and is resistant to degradation by proteases and high temperatures [191]. Cerebral accumulation of fibrillar PrP^{Sc} is diagnostic of transmissible spongiform encephalopathies (TSEs), a class of highly infectious neurodegenerative diseases for which there is no effective treatment. The extreme stability of PrP^{Sc} causes problems with medical instrument sterilisation, and TSEs can cross species barriers by ingestion of PrP^{Sc} [192].

Residues 106 to 126 of PrP are highly conserved amongst the PrP homologues, and

the synthetic PrP (106-126) peptide is amyloidogenic, neurotoxic *in vivo* and *in vitro* and protease resistant, akin to PrP^{Sc} [193]. A comparative study of PrP (106-126) and two non-amyloidogenic (*in vitro*) variants, a species where the methionine residues are oxidised [194] and a scrambled PrP (106-126) control sequence, utilised IM-MS and REMD. ATDs of PrP (106-126) and its oxidised form contained an extended conformation which was assigned by REMD secondary structure predictions, supported by CD measurements, as a β-hairpin monomer [193]. At 24 and 72 hours the ATD of the PrP (106-126) contained earlier arriving peaks, which correlated with the loss of the extended conformation. These peaks were assigned as oligomers, oligomerisation was only observed for PrP (106-126); ATDs of scrambled sequence peptide and oxidised PrP (106-126) did not change over same time period. PrP (106-126) and oxidised PrP (106-126) peptides have comparable secondary and tertiary structure, implying that the inability of methionine oxidised peptide to aggregate is a result of a reduction in inter-peptide hydrophobic interactions, which may inhibit β-sheet formation. Results from CD experiments indicate that the scrambled sequence peptide displays less β-sheet content and a higher proportion of α-helical and random coil components; structural differences which are probably accountable for its unamyloidogenic characteristics.

The conformational stability of residues 90-231 of recombinant Syrian hamster prion protein PrP, α-PrP, was studied by Grabenauer *et al.* [195]. α-PrP is α-helical, does not aggregate and is deemed to be comparable with PrP^{C}. MS shows a bimodal charge state distribution, representative of distinct conformational families present in solution. IM-MS data support this and observe an additional conformational family. Compact, intermediate and extended conformational families are observed, present in monomer charge states; +7 and +8, +7 to +11 and +12 to +15 respectively. Ions which were injected into the drift cell with low injection energy were conformationally more compact suggesting that these are the structures present in solution. Increasing the injection energy of α-PrP caused a transition to more extended conformations, with the exception of $[M+9H]^{9+}$ and $[M+10H]^{10+}$ which were exceptionally stable. Conformational stability studies, investigated using temperature, found that compact structures were converted into intermediate

structures. For $[M+7H]^{7+}$, $[M+8H]^{8+}$ and $[M+11H]^{11+}$ conversion from compact conformation to an extended state was complete by 475 K. $[M+10H]^{10+}$ completely converted to an extended structure by 575 K. The most thermally stable was $[M+9H]^{9+}$, which retained intermediate structure at 600 K. Based on the amino acid sequence of α-PrP, $[M+9H]^{9+}$ and $[M+10H]^{10+}$ are the most likely charge states to be present in solution at physiological pH, suggesting that these solution structures have an unusually high stability. The common resistance to thermal denaturation and remarkable conformational stability shared by PrP^{Sc} and α-PrP suggests features of PrP^{C} structure may be relevant to PrP^{Sc}.

Another conformational isoform of recombinant Syrian hamster prion protein PrP (90-231) was studied by IM-MS in direct comparison with α-PrP (90-231) [196]. The isoform is identical in amino acid sequence but has been refolded into a β-sheet enriched form to represent PrP^{Sc}. Contrary to α-helical α-PrP (90-231), the conformational isoform has been observed to aggregate in solution. A physiologically relevant pH study was undertaken as PrP experiences an acidic environment in the cell (pH 5.5) and a subsequent neutral pH environment when anchored to the cell membrane. At neutral pH no difference was observed in the ATDs of the two isoforms, however at pH 5.5 differences were apparent. The ATD of α-helical α-PrP (90-231) was much broader than that of PrP (90-231), suggesting the presence of multiple conformations, and contained a more significantly populated later arriving peak. The pH environment had a greater influence on α-PrP (90-231), which most closely resembles PrP^{C}, than on the β-sheet rich isoform. This study additionally advocates the use of IM-MS as a screening technique for the *in vivo* presence of possibly pathogenic protein and peptide conformational isoforms, such as PrP^{Sc}.

1.8. Summary

MS is the only technique that can separate the species present in the lag phase, identifying the many transient oligomeric orders according to their *m/z* ratios revealing their relative abundance. The hybrid technique IM-MS combined with molecular modelling enhances this picture by probing the conformational landscape of the aggregating system. In this thesis three aggregating systems are investigated; amyloidogenic transthyretin fragment (105-115), insulin and Aβ peptides.

1.9. References

1 Sipe, J. D. & Cohen, A. S. Review: History of the amyloid fibril. *J. Struct. Biol.* **130**, 88-98 (2000).

2 Sacchettini, J. C. & Kelly, J. W. Therapeutic strategies for human amyloid diseases. *Nat. Rev. Drug Discov.* **1**, 267-275 (2002).

3 Chiti, F. & Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).

4 Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678-699 (2003).

5 LeVine, H. Quantification of β-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol.* **309**, 274-284 (1999).

6 Puchtler, H., Sweat, F. & Levine, M. On Binding of Congo Red by Amyloid *J. Histochem. Cytochem.* **10**, 355 (1962).

7 Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50** 123-159 (Academic Press Inc, 1997).

8 Gazit, E. The "Correctly folded" state of proteins: Is it a metastable state. *Angew. Chem. Int. Edit.* **41**, 257 (2001).

9 Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890, (2003).

10 Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329-332 (1999).

11 Hartl, F. U. & Hayer-Hartl, M. Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858 (2002).

12 Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552-1555 (2001).

13 Ashcroft, A. E. Mass Spectrometry and the Amyloid Problem-How Far Can We Go in the Gas Phase? *J. Am. Soc. Mass. Spectrom.* **21**, 1087-1096 (2010).

14 Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).

15 Tan, S. Y. & Pepys, M. B. Amyloidosis. *Histopathology* **25**, 403-414 (1994).

16 Klein, W. L., Krafft, G. A. & Finch, C. E. Targeting small Aβ oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci.* **24**, 219-224 (2001).

17 Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267-298 (2003).

18 Kirkitadze, M. D., Bitan, G. & Teplow, D. B. Paradigm shifts in Alzheimer's disease and other neuro degenerative disorders: The emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69**, 567-577 (2002).

19 Conway, K. A. *et al.* Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *P. Natl. Acad. Sci. U. S. A.* **97**, 571-576 (2000).

20 Goldberg, M. S. & Lansbury, P. T. Is there a cause-and-effect relationship between α-synuclein fibrillization and Parkinson's disease? *Nat. Cell Biol.* **2**, 115-119 (2000).

21 Lambert, M. P. *et al.* Diffusible, nonfibrillar ligands derived from Aβ(1-42) are potent central nervous system neurotoxins. *P. Natl. Acad. Sci. U. S. A.* **95**, 6448-6453 (1998).

22 Hartley, D. M. *et al.* Protofibrillar intermediates of amyloid β-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J. Neurosci.* **19**, 8876-8884 (1999).

23 Sousa, M. M., Cardoso, I., Fernandes, R., Guimaraes, A. & Saraiva, M. J. Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy - Evidence for toxicity of nonfibrillar aggregates. *Am. J. Pathol.* **159**, 1993-2000 (2001).

24 Reixach, N., Deechongkit, S., Jiang, X., Kelly, J. W. & Buxbaum, J. N. Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *P. Natl. Acad. Sci. U. S. A.* **101**, 2817-2822 (2004).

25 Hardy, J. & Selkoe, D. J. Medicine - The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353-356 (2002).

26 Bucciantini, M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-511 (2002).

27 Kayed, R. *et al.* Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486-489 (2003).

28 Zhou, M., Ottenberg, G., Sferrazza, G. F. & Ida Lasmezas, C. Highly neurotoxic monomeric α-helical prion protein. *P. Natl. Acad. Sci. U. S. A.* **109**, 3113-3118 (2012).

29 Chapman, M. R. *et al.* Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* **295**, 851-855 (2002).

30 Berson, J. F. *et al.* Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *J. Cell Biol.* **161**, 521-533, (2003).

31 Kelly, J. W. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8**, 101-106 (1998).

32 Fandrich, M., Fletcher, M. A. & Dobson, C. M. Amyloid fibrils from muscle myoglobin - Even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature* **410**, 165-166 (2001).

33 Litvinovich, S. V. *et al.* Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J. Mol. Biol.* **280**, 245-258 (1998).

34 Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid fibril formation by an SH3 domain. *P. Natl. Acad. Sci. U. S. A.* **95**, 4224-4228 (1998).

35 Chiti, F. *et al.* Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *P. Natl. Acad. Sci. U. S. A.* **96**, 3590-3594 (1998).

36 Schmittschmitt, J. P. & Scholtz, J. M. The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Sci.* **12**, 2374-2378 (2003).

37 Channon, K. & MacPhee, C. E. Possibilities for 'smart' materials exploiting the self-assembly of polypeptides into fibrils. *Soft Matter* **4**, 647-652 (2008).

38 Mart, R. J., Osborne, R. D., Stevens, M. M. & Ulijn, R. V. Peptide-based stimuli-responsive biomaterials. *Soft Matter* **2**, 822-835 (2006).

39 Ellis-Behnke, R. G. *et al.* Nano neuro knitting: peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5054-5059 (2006).

40 Serpell, L. C. *et al.* The protofilament substructure of amyloid fibrils. *J. Mol. Biol.* **300**, 1033-1039 (2000).

41 Harper, J. D., Lieber, C. M. & Lansbury, P. T. Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-β protein. *Chem. Biol.* **4**, 951-959 (1997).

42 Blake, C. C. F. *et al.* A Molecular Model of an Amyloid Fibril in *Symposium on the Nature and Origin of Amyloid Fibrils.* 6-21 (John Wiley & Sons Ltd, Chichester, UK, 2007)

43 Goldsbury, C., Kistler, J., Aebi, U., Arvinte, T. & Cooper, G. J. S. Watching amyloid fibrils grow by time-lapse atomic force microscopy. *J. Mol. Biol.* **285**, 33-39 (1999).

44 Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M. & Griffin, R. G. Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16748-16753 (2002).

45 Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16742-16747 (2002).

46 Ritter, C. *et al.* Correlation of structural elements and infectivity of the HET-s prion. *Nature* **435**, 844-848 (2005).

47 Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 315-320 (2005).

48 Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773-778 (2005).

49 Jaroniec, C. P. *et al.* High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 711-716 (2004).

50 Channon, K. J. *et al.* Modification of fluorophore photophysics through peptide-driven self-assembly. *J. Am. Chem. Soc.* **130**, 5487-5491 (2008).

51 Deng, W., Cao, A. & Lai, L. Detecting the inter-peptide arrangement and maturation process of transthyretin (105-115) amyloid fibril using a FRET pair with short Forster distance. *Biochem. Biophys. Res. Commun.* **362**, 689-694 (2007).

52 Deng, W., Cao, A. & Lai, L. Distinguishing the cross-β spine arrangements in amyloid fibrils using FRET analysis. *Protein Sci.* **17**, 1102-1105 (2008).

53 Nazabal, A., Hornemann, S., Aguzzi, A. & Zenobi, R. Hydrogen/deuterium exchange mass spectrometry identifies two highly protected regions in

recombinant full-length prion protein amyloid fibrils. *J. Mass Spectrom.* **44**, 965-977 (2009).

54 Carulla, N. *et al.* Molecular recycling within amyloid fibrils. *Nature* **436**, 554-558 (2005).

55 Eanes, E. D. & Glenner, G. G. X-Ray Diffraction Studies on Amyloid Filaments. *J. Histochem. Cytochem.* **16**, 673 (1968).

56 Bauer, H. H. *et al.* Architecture and Polymorphism of Fibrillar Supramolecular Assemblies Produced by In-Vitro Aggregation of Human Calcitonin. *J. Struct. Biol.* **115**, 1-15 (1995).

57 Jimenez, J. L. *et al.* The protofilament structure of insulin amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9196-9201 (2002).

58 Goldsbury, C. S. *et al.* Polymorphic fibrillar assembly of human amylin. *J. Struct. Biol.* **119**, 17-27 (1997).

59 Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307**, 262-265 (2005).

60 Tanaka, M., Chien, P., Yonekura, K. & Weissman, J. S. Mechanism of cross-species prion transmission: An infectious conformation compatible with two highly divergent yeast prion proteins. *Cell* **121**, 49-62 (2005).

61 Levine, H. Thioflavine-T Interaction with Synthetic Alzheimers-Disease β-Amyloid Peptides - Detection of Amyloid Aggregation in Solution. *Protein Sci.* **2**, 404-410 (1993).

62 Naiki, H. *et al.* Establishment of a kinetic model of dialysis-related amyloid fibril extension in vitro. *Amyloid* **4**, 223-232 (1997).

63 Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317-1321 (2000).

64 Powers, E. T. & Powers, D. L. The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the "supercritical concentration". *Biophys. J.* **91**, 122-132 (2006).

65 Colaco, M., Park, J. & Blanch, H. The kinetics of aggregation of poly-glutamic acid based polypeptides. *Biophys. Chem.* **136**, 74-86 (2008).

66 Weissmann, C., Enari, M., Klohn, P. C., Rossi, D. & Flechsig, E. Transmission of Prions *Arthur M Sackler Colloquium on Self-Perpetuating Structural States in Biology, Disease and Genetics.* 16378-16383 (Natl Acad Sciences).

67 Uversky, V. N. & Fink, A. L. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *BBA-Proteins Proteom.* **1698**, 131-153, doi:10.1016/j.bbapap.2003.12.008 (2004).

68 Chiti, F. *et al.* Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo J.* **19**, 1441-1449 (2000).

69 Villegas, V. *et al.* Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci.* **9**, 1700-1708 (2000).

70 Gosal, W. S. *et al.* Competing pathways determine fibril morphology in the self-assembly of β_2-microglobulin into amyloid. *J. Mol. Biol.* **351**, 850-864 (2005).

71 Hammarstrom, P., Jiang, X., Hurshman, A. R., Powers, E. T. & Kelly, J. W. Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16427-16432 (2002).

72 Dumoulin, M. *et al.* A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature* **424**, 783-788 (2003).

73 Ray, S. S., Nowak, R. J., Brown, R. H. & Lansbury, P. T. Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3639-3644 (2005).

74 Chiti, F. *et al.* Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Arthur M Sackler Colloquium on Self-Perpetuating Structural States in Biology, Disease and Genetics.* 16419-16426 (Natl Acad Sciences).

75 Goers, J., Uversky, V. N. & Fink, A. L. Polycation-induced oligomerization and accelerated fibrillation of human α-synuclein in vitro. *Protein Sci.* **12**, 702-707 (2003).

76 Konno, T. Amyloid-induced aggregation and precipitation of soluble proteins: A electrostatic contribution of the Alzheimer's β(25-35) amyloid fibril. *Biochemistry* **40** (2001).

77 Schwartz, R., Istrail, S. & King, J. Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. *Protein Sci.* **10**, 1023-1031 (2001).

78 Broome, B. M. & Hecht, M. H. Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis. *J. Mol. Biol.* **296**, 961-968 (2000).

79 Pawar, A. P. *et al.* Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392 (2005).

80 Tartaglia, G. G. *et al.* Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).

81 Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J. M. & Brito, R. M. M. Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J. Biol. Chem.* **276**, 27207-27213 (2001).

82 Harper, J. D., Wong, S. S., Lieber, C. M. & Lansbury, P. T. Observation of metastable Aβ amyloid protofibrils by atomic force microscopy. *Chem. Biol.* **4**, 119-125 (1997).

83 Ahmed, M. *et al.* Structural conversion of neurotoxic amyloid-β1-42 oligomers to fibrils. *Nat. Struct. Mol. Biol.* **17**, 561-567 (2010).

84 Bitan, G., Lomakin, A. & Teplow, D. B. Amyloid β-protein oligomerization - Prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *J. Biol. Chem.* **276**, 35176-35184 (2001).

85 Bitan, G., Vollers, S. S. & Teplow, D. B. Elucidation of primary structure elements controlling early amyloid β-protein oligomerization. *J. Biol. Chem.* **278**, 34882-34889 (2003).

86 Modler, A. J., Gast, K., Lutsch, G. & Damaschun, G. Assembly of amyloid protofibrils via critical oligomers - A novel pathway of amyloid formation. *J. Mol. Biol.* **325**, 135-148 (2003).

87 Nguyen, H. D. & Hall, C. K. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16180-16185 (2004).

88 Bader, R., Bamford, R., Zurdo, J., Luisi, B. F. & Dobson, C. M. Probing the mechanism of amyloidogenesis through a tandem repeat of the PI3-SH3 domain suggests a generic model for protein aggregation and fibril formation. *J. Mol. Biol.* **356**, 189-208 (2006).

89 Bitan, G. *et al.* Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 330-335 (2003).

90 Harper, J. D., Wong, S. S., Lieber, C. M. & Lansbury, P. T. Assembly of Aβ amyloid protofibrils: An in vitro model for a possible early event in Alzheimer's disease. *Biochemistry* **38**, 8972-8980 (1999).

91 Smith, A. M., Jahn, T. R., Ashcroft, A. E. & Radford, S. E. Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* **364**, 9-19 (2006).

92 Krebs, M. R. H. *et al.* The formation of spherulites by amyloid fibrils of bovine insulin. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14420-14424 (2004).

93 Krebs, M. R. H., Bromley, E. H. C., Rogers, S. S. & Donald, A. M. The mechanism of amyloid spherulite formation by bovine insulin. *Biophys. J.* **88**, 2013-2021 (2005).

94 Krebs, M. R. H., Devlin, G. L. & Donald, A. M. Protein particulates: Another generic form of protein aggregation? *Biophys. J.* **92**, 1336-1342 (2007).

95 Krebs, M. R. H., Domike, K. R., Cannon, D. & Donald, A. M. Common motifs in protein self-assembly. *Faraday Discuss.* **139**, 265-274 (2008).

96 Arakawa, T., Ejima, D., Li, T. S. & Phil, J. S. The Critical Role of Mobile Phase Composition in Size Exclusion Chromatography of Protein Pharmaceuticals. *J. Pharm. Sci.* **99**, 1674-1692 (2010).

97 Philo, J. S. A Critical Review of Methods for Size Characterization of Non-Particulate Protein Aggregates. *Curr. Pharm. Biotechnol.* **10**, 359-372 (2009).

98 Andersen, C. B., Manno, M., Rischel, C., Thorolfsson, M. & Martorana, V. Aggregation of a multidomain protein: A coagulation mechanism governs aggregation of a model IgG1 antibody under weak thermal stress. *Protein Sci.* **19**, 279-290 (2010).

99 Tjernberg, L. O. *et al.* Amyloid β-peptide polymerization studied using fluorescence correlation spectroscopy. *Chem. Biol.* **6**, 53-62 (1999).

100 Wang, G., Johnson, A. J. & Kaltashov, I. A. Evaluation of Electrospray Ionization Mass Spectrometry as a Tool for Characterization of Small Soluble Protein Aggregates. *Anal. Chem.* **84**, 1718-1724 (2011).

101 Zeleny, J. Electric Discharge from Points. *Phys. Rev.* **9**, 562-563 (1917).

102 Dole, M., Mack, L. L. & Hines, R. L. Molecular Beams of Macroions. *J. Chem. Phys.* **49**, 2240 (1968).

103 Yamashita, M. & Fenn, J. B. Electrospray Ion-Source - Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **88**, 4451-4459 (1984).

104 Fenn, J. B. Electrospray wings for molecular elephants (Nobel lecture). *Angew. Chem. Int. Edit.* **42**, 3871-3894 (2003).

105 *http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/fenn-lecture.html*

106 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **246**, 64-71 (1989).
107 Ganem, B., Li, Y. T. & Henion, J. D. Observation of Noncovalent Enzyme Substrate and Enzyme Product Complexes by Ion-Spray Mass-Spectrometry. *J. Am. Chem. Soc.* **113**, 7818-7819 (1991).
108 Ganem, B., Li, Y. T. & Henion, J. D. Detection of Non-Covalent Receptor Ligand Complexes By Mass-Spectrometry. *J. Am. Chem. Soc.* **113**, 6294-6296 (1991).
109 Katta, V. & Chait, B. T. Observation of the Heme Globin Complex in Native Myoglobin by Electrospray-Ionization Mass-Spectrometry. *J. Am. Chem. Soc.* **113**, 8534-8535 (1991).
110 de Hoffmann, E., Stroobant, V. *Mass Spectrometry - Principles and Applications*. (John Wiley & Sons Ltd, Chichester, UK, 2007).
111 Tang, K. Q. & Gomez, A. Generation of Monodisperse Water Droplets from Electrosprays in a Corona-Assisted Cone-Jet Mode. *J. Colloid Interface Sci.* **175**, 326-332 (1995).
112 Iavarone, A. T. & Williams, E. R. Mechanism of charging and supercharging molecules in electrospray ionization. *J. Am. Chem. Soc.* **125**, 2319-2327 (2003).
113 Fernandez de la Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93-104 (2000).
114 Iribarne, J. V. & Thomson, B. A. Evaporation of Small Ions From Charged Droplets. *J. Chem. Phys.* **64**, 2287-2294 (1976).
115 Wilm, M. S. & Mann, M. Electrospray and Taylor-Cone Theory, Doles Beam of Macromolecules At Last *Int. J. Mass Spectrom.* **136**, 167-180 (1994).
116 Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1-8 (1996).
117 Juraschek, R., Dulcks, T. & Karas, M. Nanoelectrospray - More than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass. Spectrom.* **10**, 300-308 (1999).
118 Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: Technology for structural genomics and proteomics. *Chem. Rev.* **107**, 3544-3567 (2007).
119 Paul, W. & Steinwedel, H. Ein Neues Massenspektrometer Ohne Magnetfeld. *Z. Naturforsch. A* **8**, 448-450 (1953).
120 Stephens, W. E. A Pulsed Mass Spectrometer with Time Dispersion. *Phys. Rev.* **69**, 691-691 (1946).
121 Wiley, W. C. & McLaren, I. H. Time-of-flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **26**, 1150-1157 (1955).
122 Mamyrin, B. A., Karataev, V. I., Shmikk, D. V. & Zagulin, V. A. Mass-Reflectron a New Nonmagnetic Time-of-flight High-Resolution Mass-Spectrometer. *Zh. Eksp. I Teor. Fiz.* **64**, 82-89 (1973).
123 Comisarow, M. B. & Marshall, A. G. Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **25**, 282-283 (1974).

124 Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**, 1-35 (1998).
125 Hipple, J. A., Sommer, H. & Thomas, H. A. A Precise Method of Determining the Faraday by Magnetic Resonance *Phys. Rev.* **76**, 1877-1878 (1949).
126 Cooks, R. G., Beynon, J. H., Caprioli, R. M., and Lester, G. R. *Metastable Ions*. (Elsevier, Amsterdam, 1973).
127 Wiza, J. L. Microchannel Plate Detectors. *Nucl. Instrum. Methods* **162**, 587-601 (1979).
128 Erikson, H. A. On the effect of the medium on gas ion mobility. *Phys. Rev.* **30**, 339-348 (1927).
129 Bradbury, N. E. The absolute values of the mobility of gaseous ions in pure gases. *Phys. Rev.* **40**, 0508-0523 (1932).
130 Barnes, W. S., Martin, D. W. & McDaniel, E. W. Mass Spectrographic Identification of Ion Observed in Hydrogen Mobility Experiments. *Phys. Rev. Lett.* **6**, 110 (1961).
131 McAfee, K. B. & Edelson, D. Identification and Mobility of Ions in a Townsend Discharge by Time-Resolved Mass Spectrometry. *P. Phys. Soc. Lond.* **81**, 382 (1963).
132 Jurneczko, E. & Barran, P. E. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **136**, 20-28 (2011).
133 Pacholarz, K. J., Garlish, R. A., Taylor, R. J. & Barran, P. E. Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery. *Chem. Soc. Rev.* **41**, 4335-4355 (2012).
134 Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **43**, 1-22 (2008).
135 Harvey, S. R., MacPhee, C. E. & Barran, P. E. Ion mobility mass spectrometry for peptide analysis. *Methods* **54**, 454-461 (2011).
136 Mason, E. A., McDaniel, E. W. *Transport Properties of Ions in Gases*. (John Wiley & Sons, New York, USA, 2005).
137 Matz, L. M., Hill, H. H., Beegle, L. W. & Kanik, I. Investigation of drift gas selectivity in high resolution ion mobility spectrometry with mass spectrometry detection. *J. Am. Soc. Mass. Spectrom.* **13**, 300-307 (2002).
138 Beegle, L. W., Kanik, I., Matz, L. & Hill, H. H. Effects of drift-gas polarizability on glycine peptides in ion mobility spectrometry. *Int. J. Mass Spectrom.* **216**, 257-268 (2002).
139 Tang, K. *et al.* High-Sensitivity Ion Mobility Spectrometry/Mass Spectrometry Using Electrodynamic Ion Funnel Interfaces. *Anal. Chem.* **77**, 3330-3339 (2005).
140 Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**, 1-12 (2007).
141 Giles, K. *et al.* Applications of a travelling wave-based radio-frequencyonly stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401-2414 (2004).

142 Shvartsburg, A. A. & Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **80**, 9689-9699 (2008).
143 Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protocols* **3**, 1139-1152 (2008).
144 Bush, M. F. *et al.* Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **82**, 9557-9565 (2010).
145 Merenbloom, S., Flick, T. & Williams, E. How Hot are Your Ions in TWAVE Ion Mobility Spectrometry? *J. Am. Soc. Mass. Spectrom.* **23**, 553-562 (2012).
146 Purves, R. W., Guevremont, R., Day, S., Pipich, C. W. & Matyjaszczyk, M. S. Mass spectrometric characterization of a high-field asymmetric waveform ion mobility spectrometer. *Rev. Sci. Instrum.* **69**, 4094-4105 (1998).
147 Purves, R. W. & Guevremont, R. Electrospray ionization high-field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Anal. Chem.* **71**, 2346-2357 (1999).
148 Guevremont, R., Barnett, D. A., Purves, R. W. & Vandermey, J. Analysis of a tryptic digest of pig hemoglobin using ESI-FAIMS-MS. *Anal. Chem.* **72**, 4577-4584 (2000).
149 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86-91 (1996).
150 Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**, 16082-16086 (1996).
151 Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Mobilities of carbon cluster ions: Critical importance of the molecular attractive potential. *J. Chem. Phys.* **108**, 2416-2423 (1998).
152 Clemmer, D. E. & Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass Spectrom.* **32**, 577-592 (1997).
153 Dobo, A. & Kaltashov, I. A. Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. *Anal. Chem.* **73**, 4763-4773 (2001).
154 Kuprowski, M. C. & Konermann, L. Signal response of coexisting protein conformers in electrospray mass spectrometry. *Anal. Chem.* **79**, 2499-2506 (2007).
155 Simonson, T. & Brooks, C. L. Charge screening and the dielectric constant of proteins: Insights from molecular dynamics. *J. Am. Chem. Soc.* **118**, 8452-8458 (1996).
156 Murrell, J. N., Jenkins, A. D. *Properties of Liquids and Solutions*. 2nd edn, (John Wiley & Sons, Chichester, UK, 1994).
157 Barran, P. E. *et al.* Is it biologically relevant to measure the structures of small peptides in the gas-phase? *Int. J. Mass Spectrom.* **240**, 273-284 (2005).
158 Ouyang, Z. *et al.* Preparing protein microarrays by soft-landing of mass-selected ions. *Science* **301**, 1351-1354 (2003).

159 Blake, T. A. *et al.* Preparative linear ion trap mass spectrometer for separation and collection of purified proteins and peptides in arrays using ion soft landing. *Anal. Chem.* **76**, 6293-6305 (2004).

160 Ruotolo, B. T. *et al.* Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658-1661 (2005).

161 Antson, A. A. *et al.* Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235-242 (1999).

162 Ruotolo, B. T. & Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **10**, 402-408 (2006).

163 Breuker, K. & McLafferty, F. W. Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^{2}s. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18145-18152 (2008).

164 Hull, R. L., Westermark, G. T., Westermark, P. & Kahn, S. E. Islet amyloid: A critical entity in the pathogenesis of type 2 diabetes. *J. Clin. Endocr. Metab.* **89**, 3629-3643 (2004).

165 Westermark, P., Engstrom, U., Johnson, K. H., Westermark, G. T. & Betsholtz, C. Islet Amyloid Polypeptide - Pinpointing Amino-Acid-Residues Linked to Amyloid Fibril Formation. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5036-5040 (1990).

166 Dupuis, N. F., Wu, C., Shea, J. E. & Bowers, M. T. Human Islet Amyloid Polypeptide Monomers Form Ordered β-hairpins: A Possible Direct Amyloidogenic Precursor. *J. Am. Chem. Soc.* **131**, 18283-18292 (2009).

167 Luca, S., Yau, W. M., Leapman, R. & Tycko, R. Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. *Biochemistry* **46**, 13505-13522 (2007).

168 Dupuis, N. F., Wu, C., Shea, J.-E. & Bowers, M. T. The Amyloid Formation Mechanism in Human IAPP: Dimers Have β-Strand Monomer-Monomer Interfaces. *J. Am. Chem. Soc.* **133**, 7240-7243 (2011).

169 Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. β_2-Microglobulin - A Free Immunoglobulin Domain. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1697-1701 (1972).

170 Ljunggren, H. G. *et al.* Empty MHC Class-1 Molecules Come Out in the Cold. *Nature* **346**, 476-480 (1990).

171 Bach, M. L., Huang, S. W., Hong, R. & Poulik, M. D. β_2-Microglobulin - Association with Lymphocyte Receptors. *Science* **182**, 1350-1352, doi:10.1126/science.182.4119.1350 (1973).

172 Gejyo, F. *et al.* A New Form of Amyloid Protein Associated with Chronic-Hemodialysis was Identified as β_2-Microglobulin. *Biochem. Biophys. Res. Commun.* **129**, 701-706 (1985).

173 Chowdhury, S. K., Katta, V. & Chait, B. T. Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **112**, 9012-9013 (1990).

174 Borysik, A. J., Radford, S. E. & Ashcroft, A. E. Co-populated conformational ensembles of β_2-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *J. Biol. Chem.* **279**, 27069-27077 (2004).

175 Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E. & Ashcroft, A. E. Monitoring copopulated conformational states during protein folding events using Electrospray ionization-ion mobility spectrometry-mass spectrometry. *J. Am. Soc. Mass Spectr.* **8**, 2180-2190 (2007).

176 McParland, V. J. *et al.* Partially unfolded states of β_2-microglobulin and amyloid formation *in vitro*. *Biochemistry* **39**, 8735-8746 (2000).
177 Hoaglund-Hyzer, C. S., Counterman, A. E. & Clemmer, D. E. Anhydrous protein ions. *Chem. Rev.* **99**, 3037-3079 (1999).
178 Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**, 1-12 (2007).
179 McParland, V. J. *et al.* Partially unfolded states of β_2-microglobulin and amyloid formation in vitro. *Biochemistry* **39**, 8735-8746 (2000).
180 Smith, D. P., Radford, S. E. & Ashcroft, A. E. Elongated oligomers in β_2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6794-6798 (2010).
181 Smith, D. P., Woods, L. A., Radford, S. E. & Ashcroft, A. E. Structure and Dynamics of Oligomeric Intermediates in β_2-Microglobulin Self-Assembly. *Biophys. J.* **101**, 1238-1247 (2011).
182 Spillantini, M. G. *et al.* α-synuclein in Lewy bodies. *Nature* **388**, 839-840 (1997).
183 Volles, M. J. *et al.* Vesicle permeabilization by protofibrillar α-synuclein: Implications for the pathogenesis and treatment of Parkinson's disease. *Biochemistry* **40**, 7812-7819 (2001).
184 Frimpong, A. K., Abzatimov, R. R., Uversky, V. N. & Kaltashov, I. A. Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: Conformational heterogeneity of α-synuclein. *Proteins* **78**, 714-722 (2010).
185 Ly, T. & Julian, R. R. Protein-Metal Interactions of Calmodulin and α-Synuclein Monitored by Selective Noncovalent Adduct Protein Probing Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **19**, 1663-1672 (2008).
186 Uversky, V. N. A protein-chameleon: Conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders. *J. Biomol. Struct. Dyn.* **21**, 211-234 (2003).
187 Bernstein, S. L. *et al.* α-Synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation. *J. Am. Soc. Mass. Spectrom.* **15**, 1435-1443 (2004).
188 Uversky, V. N., Li, J. & Fink, A. L. Evidence for a partially folded intermediate in α-synuclein fibril formation. *J. Biol. Chem.* **276**, 10737-10744 (2001).
189 Grabenauer, M. *et al.* Spermine binding to Parkinson's protein α-synuclein and its disease-related A30P and A53T mutants. *J. Phys. Chem. B* **112**, 11147-11154 (2008).
190 Pan, K. M. *et al.* Conversion of α-Helices into β-Sheets Features in the Formation of the Scrapie Prion Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10962-10966 (1993).
191 Fernie, K., Steele, P. J., Taylor, D. M. & Somerville, R. A. Comparative studies on the thermostability of five strains of transmissible-spongiform-encephalopathy agent. *Biotechnol. Appl. Biochem.* **47**, 175-183 (2007).
192 Collinge, J. Variant Creutzfeldt-Jakob disease. *Lancet* **354**, 317-323 (1999).

193 Grabenauer, M., Wu, C., Soto, P., Shea, J.-E. & Bowers, M. T. Oligomers of the Prion Protein Fragment 106-126 Are Likely Assembled from β-Hairpins in Solution, and Methionine Oxidation Inhibits Assembly without Altering the Peptide's Monomeric Conformation. *J. Am. Chem. Soc.* **132**, 532-539 (2010).

194 Heegaard, P. M. H., Pedersen, H. G., Flink, J. & Boas, U. Amyloid aggregates of the prion peptide PrP106-126 are destabilised by oxidation and by the action of dendrimers. *FEBS Lett.* **577**, 127-133 (2004).

195 Grabenauer, M. *et al.* Conformational Stability of Syrian Hamster Prion Protein PrP(90-231). *J. Am. Chem. Soc.* **132**, 8816 (2010).

196 Hilton, G. R. *et al.* Structural Analysis of Prion Proteins by Means of Drift Cell and Traveling Wave Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **21**, 845-854 (2010).

2

Experimental and Method Development

Descriptions of the instruments used in the following chapters are detailed, with schematic representations and typical tuning parameters. An example of data acquisition and processing is given for an ion mobility mass spectrometry experiment. Nano-electrospray conditions and solution preparation details are also provided.

2.1. Reagents

Organic solvents were analytical HPLC grade and supplied by either Sigma Aldrich (Dorset, UK) or Fisher Scientific Ltd (Loughborough, UK). Acids were supplied by VWR International Ltd (UK). Very pure water was distilled using an Arium 611 water purification unit (Sartorius, Göttingen, Germany) or supplied by Fisher Scientific Ltd (Loughborough, UK). Solids were weighed with a Mettler AC100 balance to an accuracy of ±0.0001 g. Solution pH readings were taken using a Jenway 3505 pH meter to an accuracy of 0.01 units. Details of peptide and protein suppliers and purities can be found in the relevant chapters. Sodium iodide, used as a calibrant, was supplied by Sigma Aldrich (Dorset, UK).

2.2. Mass Spectrometry

2.2.1. Sample Introduction - Nano-Electrospray Ionisation

All MS and IM-MS instruments utilised were fitted with nano-electrospray (n-ESI) sources. n-ESI capillaries were prepared in house from thin walled glass capillaries (Precision Instruments, Stevenage, UK) using a micropipette puller (Fleming/Brown P-97, Sutter Instruments, Novato, CA, USA). Sample solutions were inserted into capillaries using gel micro-loading tips (Eppendorf, Hamburgh, Germany) and charged *via* the insertion of a thin platinum wire, to which a voltage was applied. Capillary voltage was adjusted to give optimum spraying conditions and ideally a consistent visible plume of ions. The precise voltage used depended on the sample solution composition (e.g. high organic content required low capillary voltages due to increased solution volatility) but voltages were normally in the region of 1.6 kV. For viscous aggregated samples a nitrogen backing gas was applied to encourage a good spray.

Application of decreasing voltages diverts the plume of nebulised ions, produced by the n-ESI source, perpendicularly into the mass spectrometer through the sample cone opening, and again perpendicularly through another source component called the extractor cone (Figure 2.2). This is known as a Z Spray™ ion source, and has the advantage of eliminating the bulk of the neutral species (excess solvent, uncharged proteins etc), thus preventing the need for frequent cleaning of the source components. Unless stated otherwise the source block was heated to 80 $^{\circ}$C to promote ion desolvation. Ions then pass through the RF lens and into the quadrupole analyser.

2.2.1.1. Elevating the Source Pressure

Gentle desolvation and an elevated source pressure enabled the transmission of larger aggregates with minimal fragmentation. To achieve elevated source pressures a Speedivalve (SP25, Edwards, West Sussex, UK) was inserted into the vacuum line between the source rotary backing pump and the source block. Partially closing the Speedivalve, thus throttling the source backing pump, enabled source backing pressures of up to one mBar (as measured by a Pirani gauge (Edwards, West Sussex, UK) on the source rotary) to be achieved (Figure 2.3). All QTOF MS instruments and the IM-MS instrument utilised in the following experimental chapters possessed this source modification. Optimisation of the source pressure was undertaken for each system studied, the most favourable pressure being approximately $3x10^{-1}$ mBar for all instruments.

2.2.1.2. Source Heating Block

To enable samples to be heated during an MS or IM-MS experiment, a source heating block was designed and constructed. This consists of an aluminium block to hold the capillary carriage, with an additional space for the insertion of a cartridge heater. The cartridge heater used was rated 120 V and 20 W from the CSS series (Omega, USA). A power pack applies voltage to the cartridge heater, which is

regulated by a K type thermocouple in a thermostatic feedback loop enabling the aluminium block to be heated to a set temperature. When the desired temperature is reached, the power pack ceases to supply a voltage and subsequently, when the temperature is detected to have dropped below the set value, a voltage is reapplied.

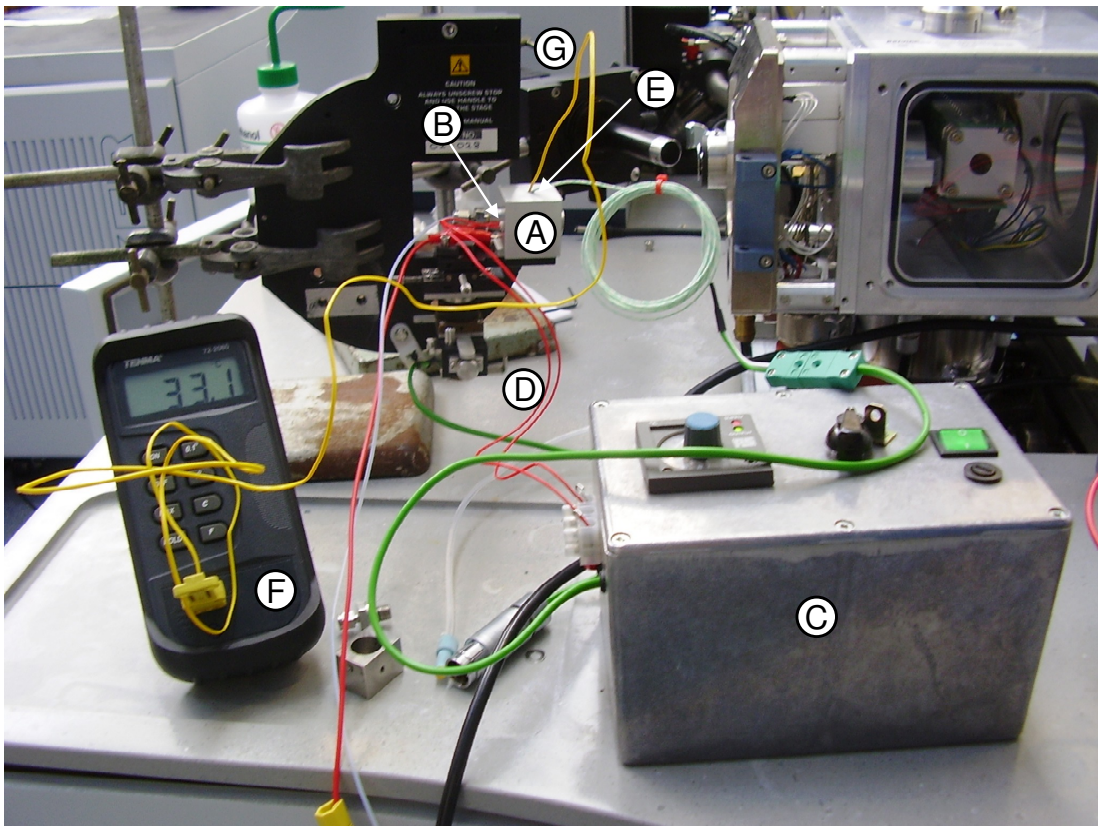


Figure 2.1 | Photo of the modified source block (A) with cartridge heater (B) inserted. The power supply (C) supplies a voltage via the red wires (D) to the cartridge heater. A thermocouple (E) detects the temperature of the source block and inputs this information via the green wires to the power pack (C). The temperature of the source block is detected by a thermocouple (G) and externally verified with a portable thermometer (F).

2.2.2. Ion Transfer, Analysis and Detection

The quadrupole analyser can be used either as a mass filter or to select only ions of a specific *m/z* value for collision induced dissociation (CID). CID is performed in the hexapole collision cell (Figure 2.2), filled with argon gas at an adjustable pressure of ~1.5 mBar. When CID is not performed ions traverse the collision cell after being accelerated through a potential difference, and the presence of argon gas improves the resolution of the resulting spectra. When CID is performed, ions of a specific *m/z* will enter the collision cell at increasing kinetic energies. Fragmentation of the ions will eventually occur and these fragments are recorded in the TOF detector.

Upon exiting the collision cell ions pass through a transfer hexapole and into the TOF region (Figure 2.2). The pusher accelerates the ions into the TOF tube until they reach the reflectron where they are refocused and reflected back to the microchannel plate (MCP) detector. Ion detection signals are converted by a 4 GHz time-to-digital converter in the embedded computer and relayed to MassLynx™ software, resulting in a total ion count (TIC) chromatogram which can be deconvoluted into a mass spectrum.

2.2.3. Instruments

Instruments used were a QTOF 2 and a QTOF Ultima, both manufactured by Waters (Manchester, UK).

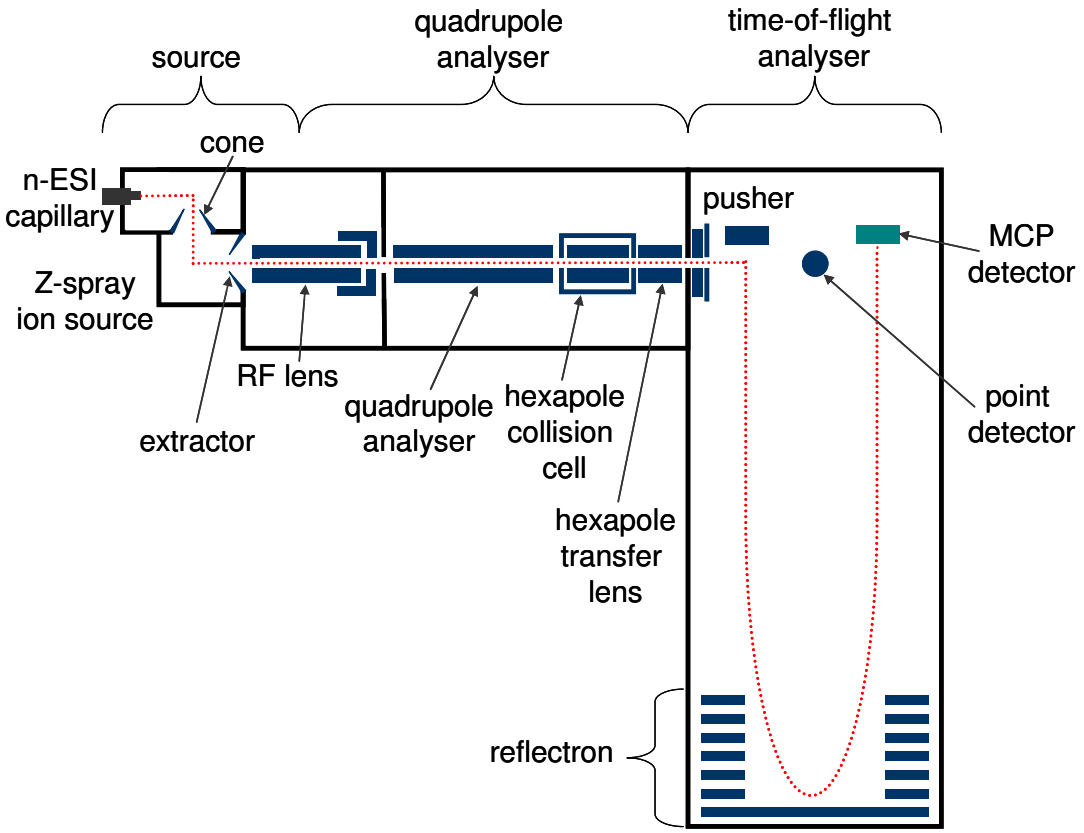


Figure 2.2 | Schematic of a QTOF mass spectrometer. All relevant components are labelled.

QTOF mass spectrometers have three sections; the source, quadrupole and TOF, which are pumped differentially to create a decreasing pressure gradient along the length of the mass spectrometer (Figure 2.2). Each region is kept at vacuum by a turbomolecular pump, backed by a rotary pump (Figure 2.3). Representative pressures for each region of the instruments used are detailed in Table 2.1, Table 2.2 and Table 2.3.

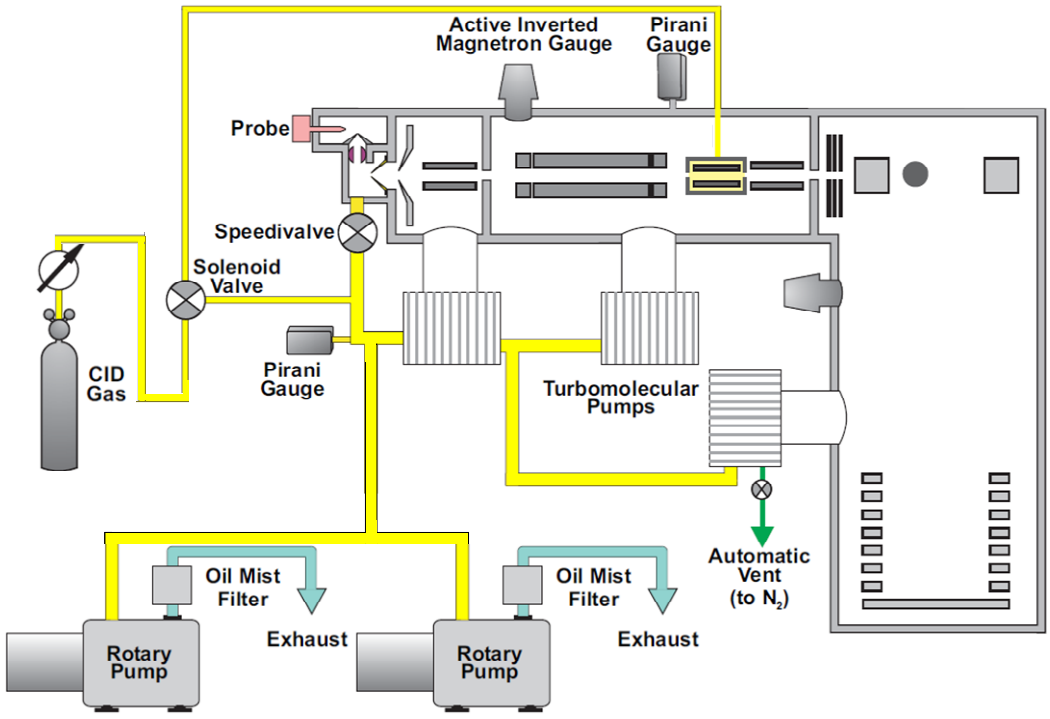


Figure 2.3 | Schematic of a QTOF mass spectrometer with pumping components and vacuum gauges labelled. The schematic is adapted from the Waters QTOF 2 users manual.

All instruments were calibrated with a 2 $mgml^{-1}$ aqueous solution of NaI, with known *m/z* values of NaI clusters used to obtain an accurate calibration. Data were processed using MassLynx™ Software 4.1 (Waters, Manchester, UK).

2.2.3.1. Instrument Tuning Parameters

Instrument Parameters	Setting
Capillary / kV	1 to 2
Cone / V	10 to 40
RF Lens	0.5
Aperture 1	0
Source / ℃	80
LM Resolution	5.0
HM Resolution	5.0
Collision Energy	8.0
Ion Energy	2.0
Steering / V	0.00
Entrance / V	70.0
Pre-filter	6.0
Transport	2.0
Aperture 3	9.0
Acceleration / V	200
Focus / V	0
Tube Lens / V	100
Offset 1 / V	-0.3
Offset 2 / V	0.0
Pusher / V	980
TOF / kV	9.10
Reflectron	35.52
Pusher Cycle Time / µs	150
Pusher Frequency / Hz	666.6667
Multiplier	550
MCP	2000
TDC Start / mV	750.0
TDC Stop / mV	150.0
TDC Threshold	0.0
Source Pressure / mBar	3.07×10^{-1}
Analyser Pressure / mBar	1.45×10^{-3}
TOF Pressure / mBar	4.27×10^{-7}

Table 2.1 | Typical Settings for the QTOF Ultima in positive polarity.

Instrument Parameters	Settings / +ve mode	Settings / -ve mode
Capillary / kV	1 to 2	1 to 2
Cone / V	10 to 40	10 to 40
Extractor	0	0
RF Lens	0.55	0.55
Source / ℃	80	80
LM Resolution	5.0	5.0
HM Resolution	5.0	5.0
Collision Energy	10.0	10.0
Ion Energy	2.7	2.8
Steering / V	0.00	1.00
Entrance / V	70.0	65.0
Pre-filter	6.0	5.0
Transport	3.7	20.0
Aperture 2	8.0	14.5
Acceleration / V	200	200
Focus / V	1	0
Tube Lens / V	100	120
Offset 1 / V	-0.5	0.0
Offset 2 / V	0.0	0.0
Pusher / V	980	980
TOF / kV	9.1	9.1
Reflectron	34.66	34.66
Pusher Cycle Time / µs	124	124
Pusher Frequency / Hz	8064.5161	8064.5161
Multiplier	650	550
MCP	2000	2000
TDC Start / mV	300.0	300.0
TDC Stop / mV	50.0	50.0
TDC Threshold	0.0	0.0
Source Pressure / mBar	1.75×10^{0}	1.77×10^{0}
Analyser Pressure / mBar	1.5×10^{-3}	1.5×10^{-3}
TOF Pressure / mBar	3.39×10^{-7}	3.39×10^{-7}

Table 2.2 | Typical Settings for the QTOF 2 in positive and negative polarities.

2.3. Ion Mobility Mass Spectrometry

2.3.1. Instrument

Ion mobility mass spectrometry measurements were performed on an in-house modified QToF 1 (Waters, Manchester, UK), known as the "MoQTOF" [1]. It has been adapted to be capable of making temperature dependent CCS measurements *via* the inclusion of a 5.1 cm long copper drift cell and supplementary ion optics, which are situated post source optics and prior to the quadrupole analyser. The drift cell is filled with helium gas (BOC Speciality Gases, purity 99.999 %) at a pressure of between 3.0 and 3.7 Torr and a weak static electric field is applied to compel ions to drift through. The pressure in the drift cell is measured with a Baratron (MKS Instruments, Andover, Massachusetts, USA). Although the drift cell has heating and cooling capacity, all measurements presented in this thesis were taken at ambient temperature (~300 K). As a consequence of the inclusion of a drift cell, and to avoid overloading the standard turbomolecular pumps, the drift cell chamber is kept at vacuum by the inclusion of an extra turbomolecular pump, backed by a dual stage rotary pump (Edwards, West Sussex, UK).

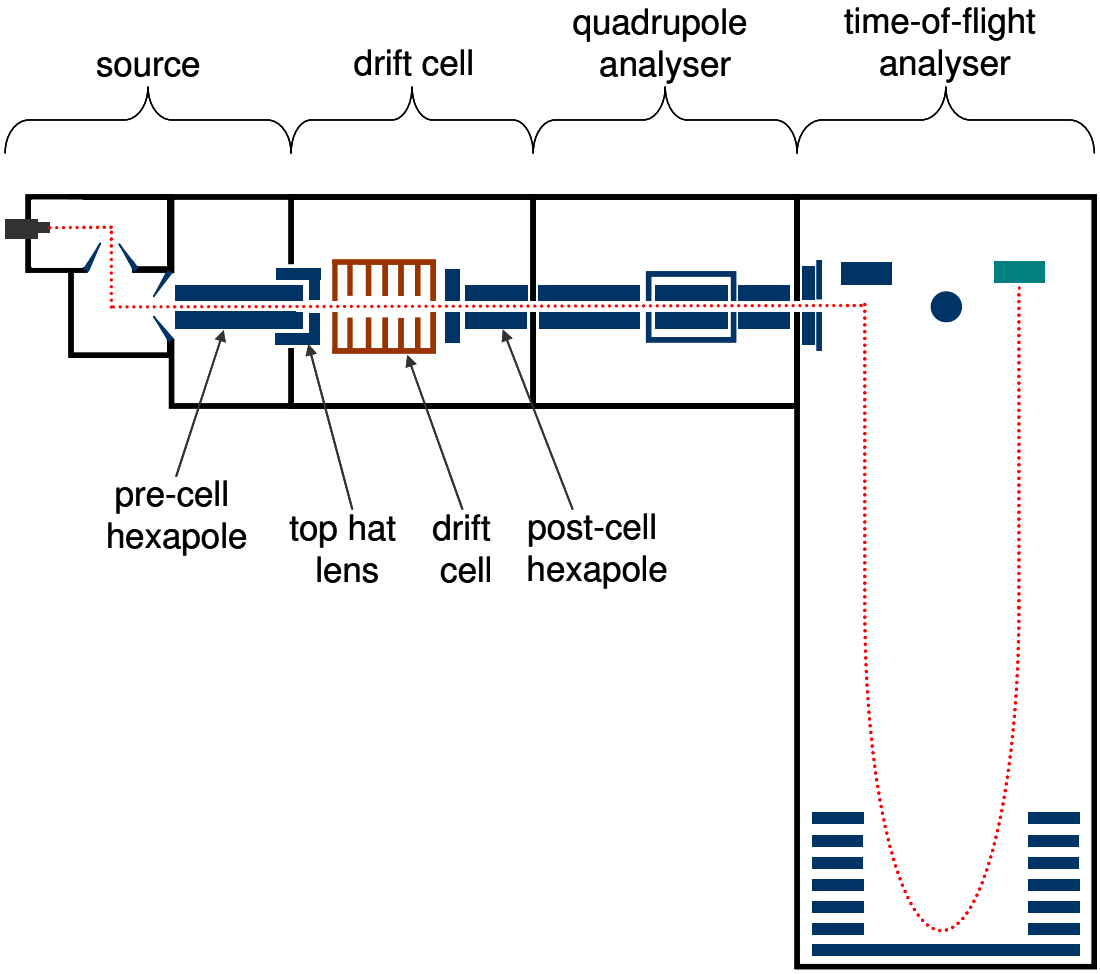


Figure 2.4 | Schematic of the drift tube IM-MS instrument – the "MoQTOF". In house modifications are labelled; all other components are as labelled in Figure 2.2.

The MoQTOF can be used in both MS and IM-MS modes; in the former ions drift continuously through the cell, whereas in the latter ions are released into the drift cell in discrete pulses. These pulses are achieved by applying a 'trapping voltage' to the top hat lens which leads to the accumulation of ions at the entrance to the drift cell (Figure 2.4). Periodically the voltage on the top hat lens is lowered for a duration of 40 µs and the accumulated packet of ions enters the drift cell. The pusher period of the TOF, itself determined by the *m/z* range required to observe all experimentally interesting species, determines the frequency of ion packets entering the cell. The frequency is regulated by a Stanford generator (Stanford Research Systems, Sunnyvale, CA, USA).

On exiting the drift cell ions are refocused in the post-cell hexapole, pass through a quadrupole mass analyser which is run as a wide band pass filter, and finally travel into a time of flight analyser where they are detected by the MCP detectors. Arrival time distributions are recorded for all of the ions detectable, and may be deconvoluted into selected ion arrival time distributions (ATDs) *via* MassLynx™ 4.1 (Waters, Manchester, UK).

2.3.1.1. Instrument Tuning Parameters

Instrument Parameters	Settings / +ve mode	Settings / -ve mode
Capillary / kV	1 to 2	1 to 2
Cone / V	10 to 40	10 to 90
Extractor	110	110
RF Lens	0.55	0.55
Source / ℃	80	55 – 80
LM Resolution	5.0	5.0
HM Resolution	5.0	5.0
Collision Energy	2.3	2.3
Ion Energy	2.0	2.0
Steering / V	0.00	0.00
Entrance / V	39.9	39.9
Pre-filter	10.8	10.8
Transport	5.0	5.0
Aperture 2	12.6	12.6
Acceleration / V	200	200
Focus / V	1	0
Tube Lens / V	80	80
Guard	64.0	64.0
TOF / kV	7.20	7.20
Reflectron	35.0	35.0
Pusher Cycle Time / µs	105	105
Pusher Frequency / Hz	9523.8095	9523.8095
Multiplier	655	655
MCP	3000	3000
TDC Start / mV	300.0	300.0
TDC Stop / mV	40.0	40.0
TDC Threshold	0.0	0.0
Source Pressure / mBar	1.4×10^{-1}	1.4×10^{-1}
Analyser Pressure / mBar	1.45×10^{-3}	1.36×10^{-3}
TOF Pressure / mBar	2.0×10^{-7}	1.2×10^{-7}

Table 2.3 | Typical Settings for the MoQTOF in positive and negative polarities.

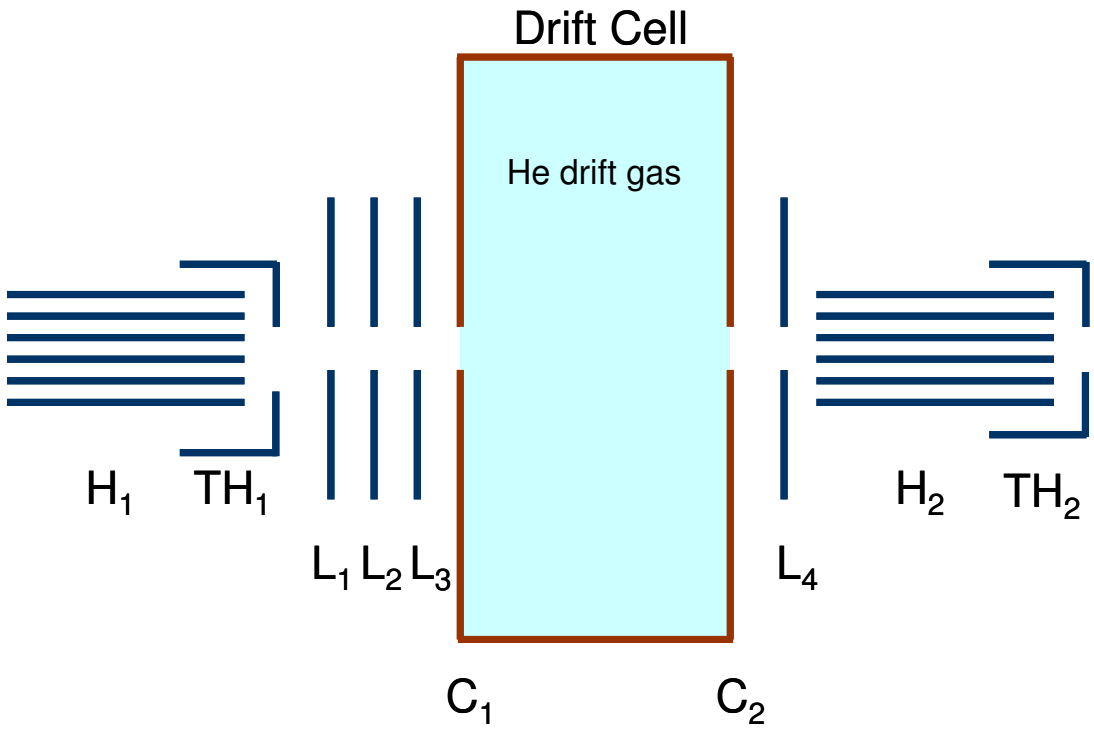


Figure 2.5 | Schematic of the MoQTOF drift cell with surrounding lenses. H_1 is the pre-cell hexapole, H_2 the post-cell hexapole, TH1 and TH2 pre and post cell top hat lenses. L_1, L_2 and L_3 pre-cell lenses, L_4 post-cell lens. C_1 cell body one and C_2, cell body two.

The voltages applied to the additional lenses and drift cell, required for the MoQTOF to have ion mobility capability, are generated by an external power supply. Pre drift cell lenses (H_1, TH_1, L_1, L_2 and L_3) are referenced to C_1 and the post drift cell lens (L_4) is referenced to C_2. H_2 and TH_2 both float above the collision cell voltage (CV) which is set by the instrument tuning page and C_2 and H_1 are referenced to this and the source voltage respectively, which is set manually by the internal power supply. A schematic of the drift cell with surrounding lenses is shown in Figure 2.5 and representative drift cell lens voltage settings are tabulated in Table 2.4.

Lens	Applied Voltage / V	Actual Voltage / V
H_1	35	$H_1 + C_1 = 105$
TH_1	7	$TH_1 + C_1 = 77$
L_1	-27	$L_1 + C_1 = 43$
L_2	-100	$L_2 + C_1 = 30$
L_3	21	$L_3 + C_1 = 91$
C_1	70	Ref CV
C_2	10	Ref CV
L_4	-8	$L_4 + C_2 = 2$

Table 2.4 | Typical Drift Cell Lens Stack Settings for the MoQTOF for IM-MS in positive mode.

2.3.1.2. Injection Energy Experiments

The kinetic energy with which ions enter the drift cell can be increased by raising the potential difference between the pre-cell hexapole and the first drift cell lens. An injection voltage range between 25 V to 54 V can be obtained (at present restricted by power supply limitations). Individual experimental chapters detail the specific injection energies used.

2.3.2. Example IM-MS Experiment

The following experiment illustrates the method of IM-MS data acquisition and analysis.

An ion which adopts an extended conformation will take longer to traverse the drift cell, as a result of undergoing more collisions with the buffer gas, than a smaller, more compact, structure. This allows disparate structures to be distinguished which are identical in mass and chemical composition. This has previously been

demonstrated for synthetic cyclic and linear peptides [2], however it can also be used to separate organic compounds.

For example, the ligand $C_{89}H_{84}N_8O_6$ can be knotted by the action of metal ions, producing an unknot macrocycle and a trefoil knot (Figure 2.6) [3].

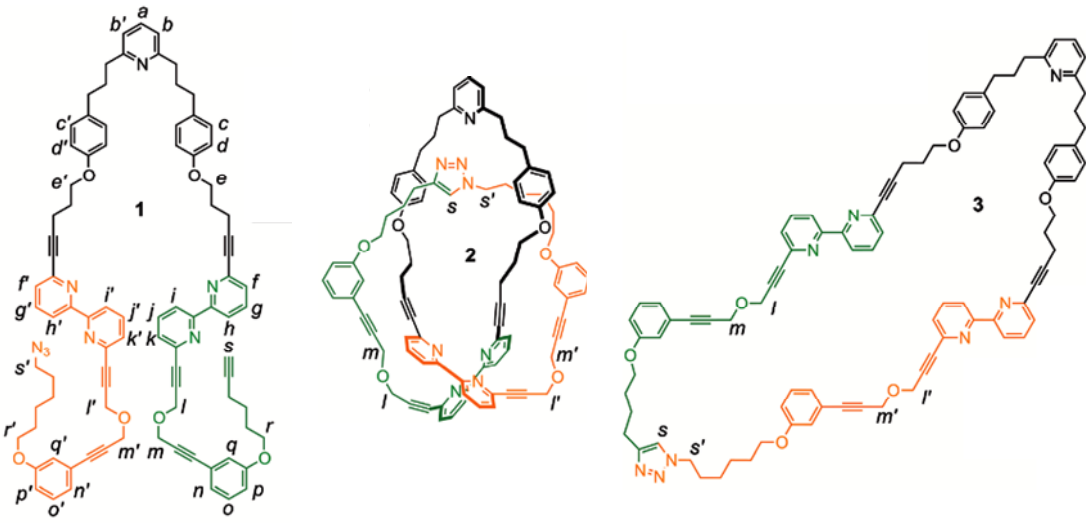


Figure 2.6 | Chemical structures of ligand isomers. (1) precursor ligand (2) trefoil knot and (3) unknot macrocycle. Figure adapted from Barran et al. [3]

These three isomers all have the same molecular formula, but can be distinguished by their size and the relative flexibility of the multiply charged molecular ions produced by the n-ESI process. Under acidified conditions the following mass spectra are produced:

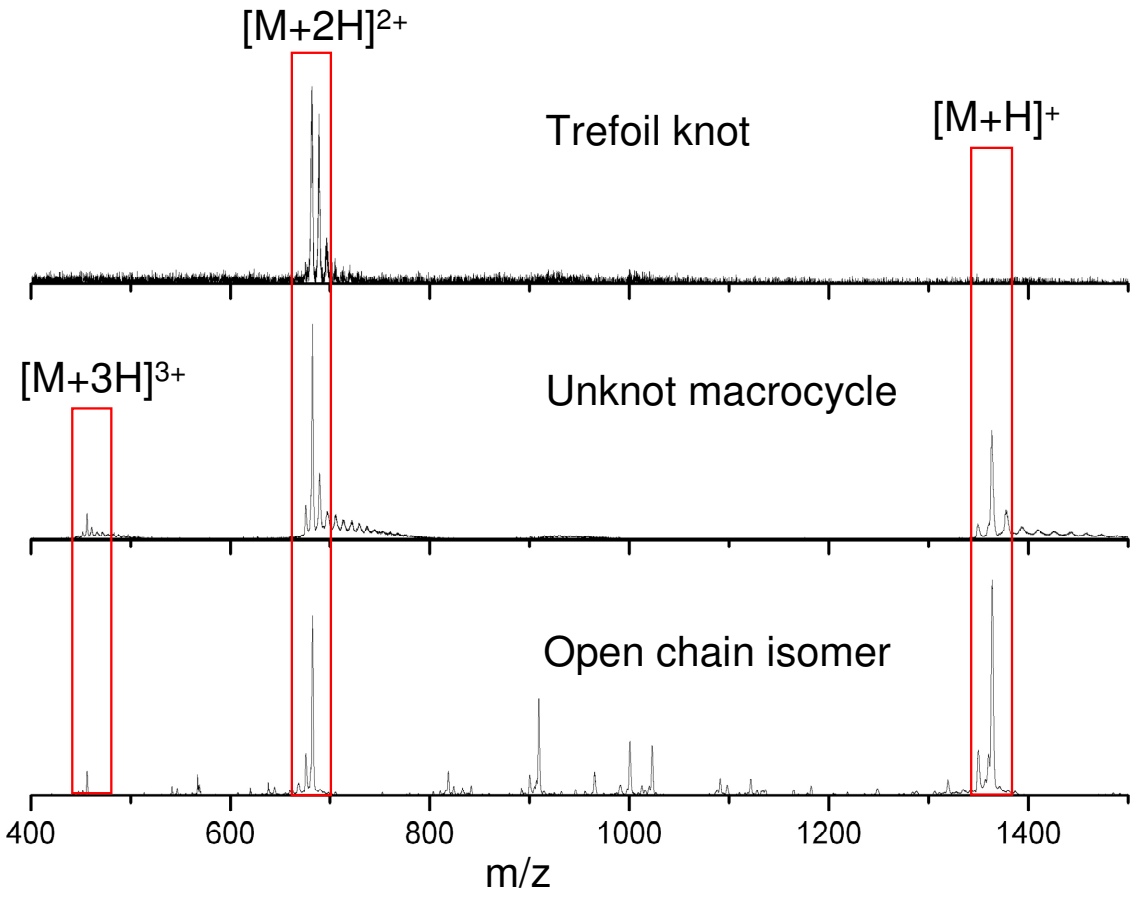


Figure 2.7 | Mass spectra of the three compounds acquired on the MoQTOF, unassigned peaks in the open chain isomer mass spectrum correspond to fragments.

For all species the $[M+2H]^{2+}$ ion is the most dominant charge state present. In the case of the trefoil knot this is the only significant species present, suggesting a very conformationally restricted structure. The absence of the $[M+3H]^{3+}$ species for the trefoil knot is indicative of a compact conformation unable to support the mutual electrostatic repulsions of three protons. IM-MS can be used to quantify these observations.

IM-MS data were taken for each isomer at eight drift voltages; 60 V, 50 V, 40 V, 35 V, 30 V, 25 V, 20 V and 15 V. For each drift voltage a total ion ATD is generated

every 200 scans, until at least 12 were collected (Figure 2.8). From the total ion ATD a mass spectrum can be generated (Figure 2.7) and from this a deconvoluted ATD of an individual *m/z* peak created.

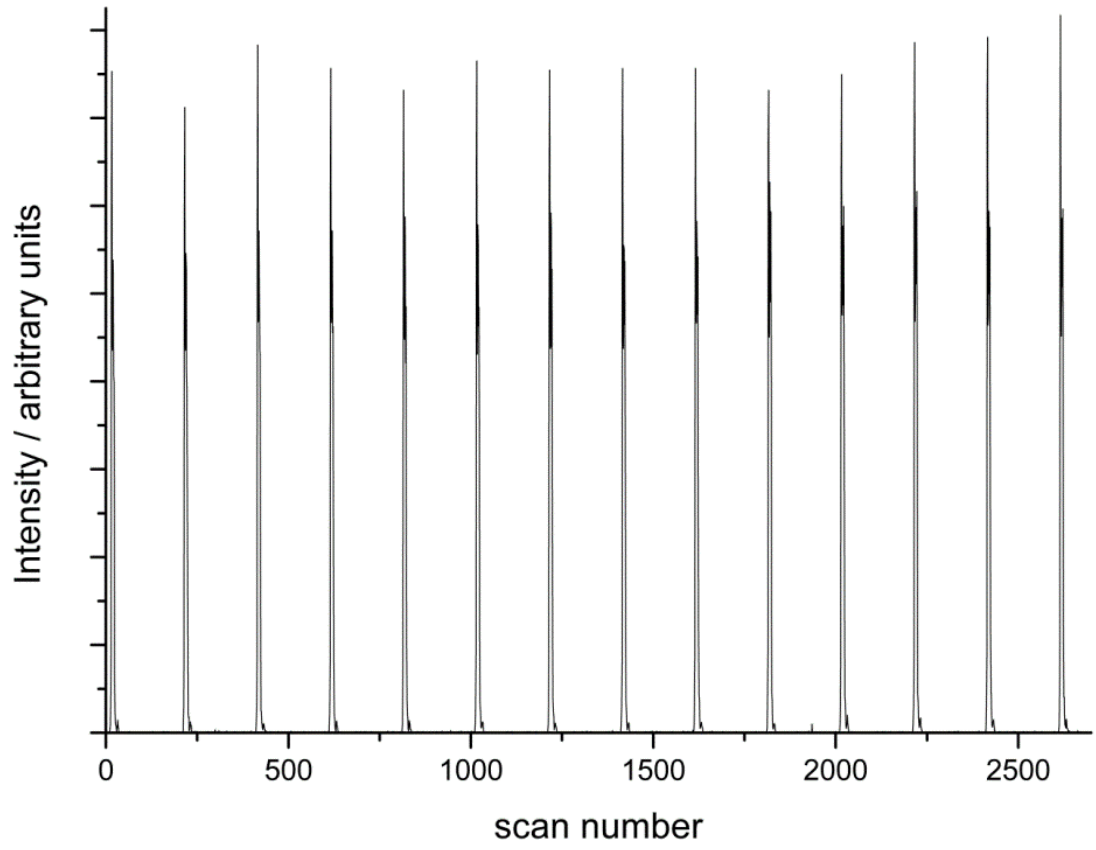


Figure 2.8 | Trefoil knot total ion ATD at a drift voltage of 50 V. Each peak is a collation of 200 scans.

Deconvoluted ATD peaks are summed and the average scan number of the resulting distribution multiplied by the pusher period to give an arrival time for the selected species at this drift voltage. Arrival time, t_a, measured by the ATD, is the time from ion injection into the cell until detection. It equates to the time taken for the ions to drift through the cell, t_d, plus the time the ions spent outside the drift cell (known as the dead time t_0). The arrival time of ions results from their low field mobility (K), which is inversely related to their rotationally averaged CCS (Ω) (Chapter 1, Section 1.5.11, Equation 1.14).

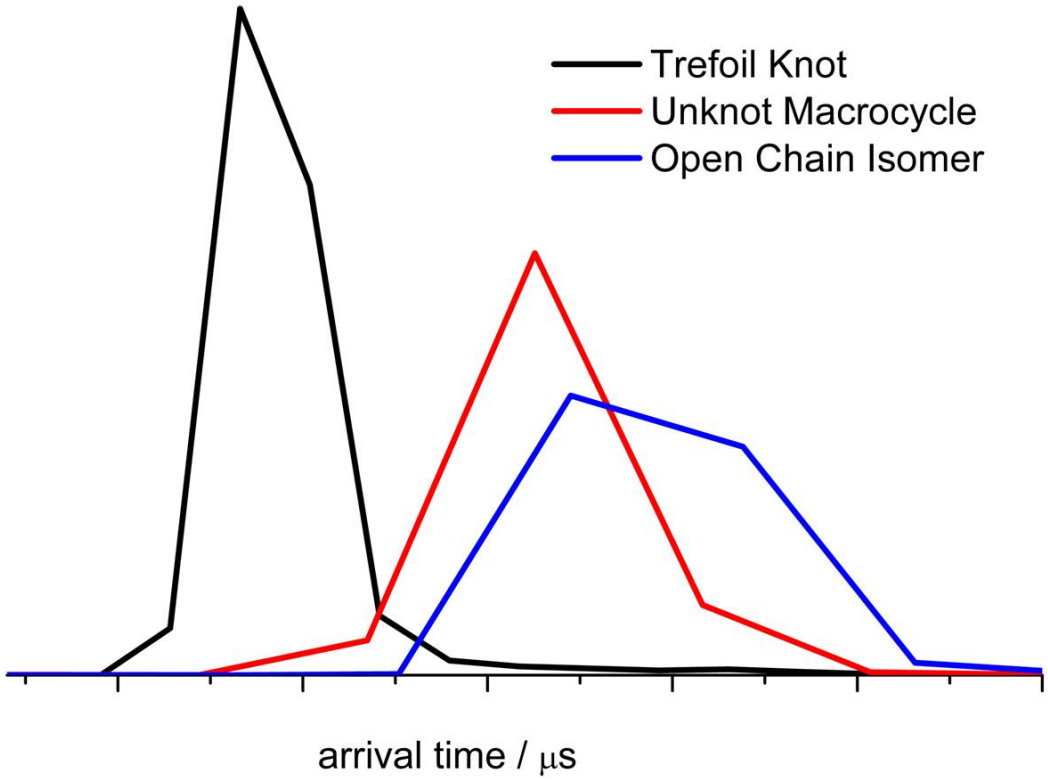


Figure 2.9 | Comparison of highest observed charge state ATDs of the trefoil knot $[M+2H]^{2+}$, unknot macrocycle $[M+3H]^{3+}$ and the open chain isomer $[M+3H]^{3+}$. Data shows the ATDs at a drift voltage of 50 V. Intensities are normalised to the peak areas.

A graph of experimentally measured arrival times at different drift voltages against *P/V* will give a linear relationship with an intercept equal to t_0 and slope of gradient *1/K* (Figure 2.10). The pressure and temperature of the helium gas are recorded at the start and end of each mobility experiment, the average of each taken and these values (*P* and *T*) used to convert *K* into reduced mobility K_0. The equation for the time spent by the ions in the drift cell, t_d, can therefore be calculated: *P* and *T* are averaged experimental helium pressure and temperature, *V* is the voltage applied across the drift cell, *L* is the length of the drift cell, K_0 is reduced low field mobility, T_0 is normalised temperature (273.15 K) and P_0 is normalised pressure (760 Torr).

$$t_d = t_a - t_0 = \frac{P}{V}\frac{L^2 T_0}{K_0 P_0 T} \qquad \textit{Equation 2.1}$$

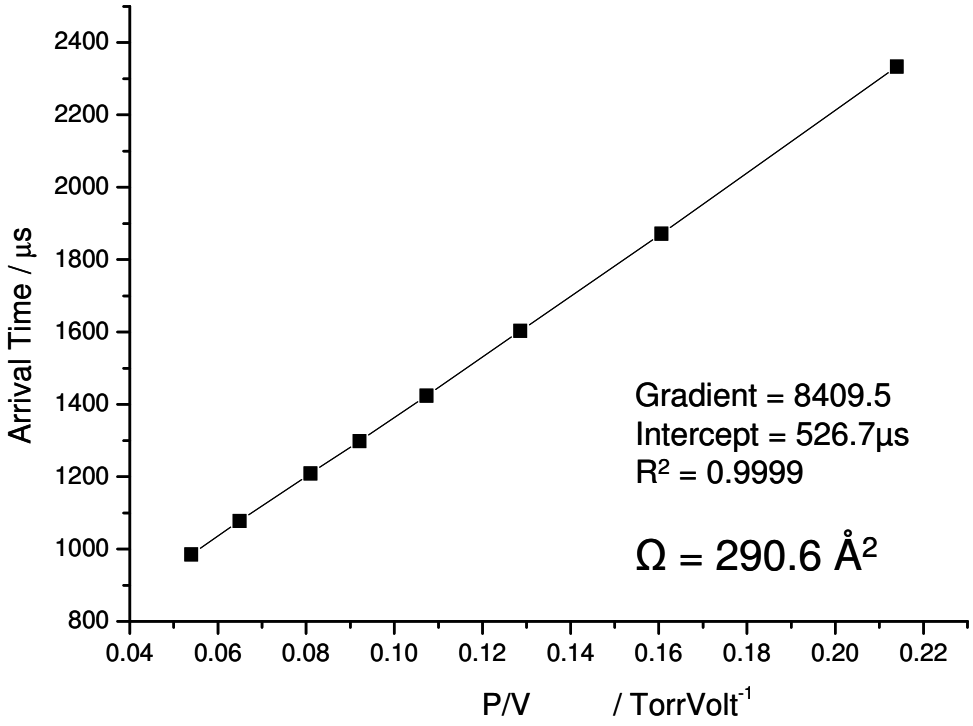


Figure 2.10 | Graph of P/V versus arrival time for the trefoil knot $[M+2H]^{2+}$ species.

The gradient of the line of best fit is equal to the inverse of the reduced low field ion mobility. The linear fit should have a high R^2 value (> 0.9995, and ideally = 0.9999), deviations from this imply either low quality experimental data, or, if deviation is observed at the higher drift voltages, that the experiment was not conducted in the low field limit. If the fit is good then the gradient can be used to calculate the rotationally averaged CCS of the molecular species using Equation 2.2. Ω is the momentum transfer collision integral (rotationally averaged CCS), z is nominal ion charge, e is elementary charge, N_0 is the buffer gas number density, μ is the reduced mass of the buffer gas and ion, k_b is the Boltzmann constant, T is the effective temperature and K_0 is reduced low field mobility.

$$\Omega = \frac{3ze}{16N_0}\left(\frac{2\pi}{\mu k_b T}\right)^{\frac{1}{2}}\frac{1}{K_0} \qquad \textit{Equation 2.2}$$

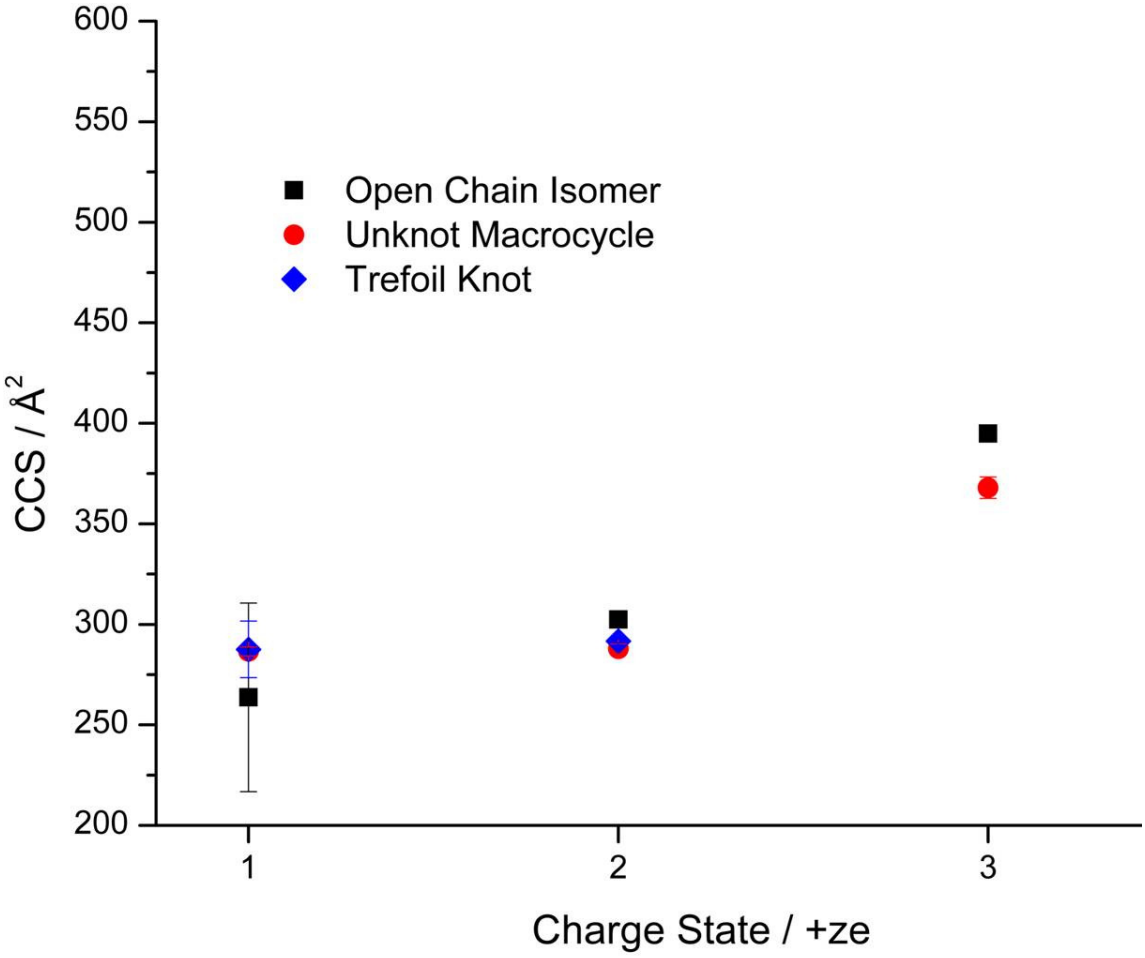


Figure 2.11 | CCSs for all observed species of the three isomers. Data are the average of three repeats; error bars represent the associated standard deviation.

As the number of charges on a molecular ion increases, the size of the adopted conformations increases as electrostatic repulsions try to force the largest distance between the charges that the molecule will allow. In general, observed CCSs increase with the amount of charge that a flexible structure carries. In addition to the magnitude of the CCS, the broadness of the distribution indicates the flexibility (the number of differently sized and shaped conformations adopted) of the molecular structure.

The open chain isomer, which is expected to have a large degree of flexibility, exhibits the broadest ATD. The unknot macrocycle has a narrower ATD and the

trefoil knot displays the narrowest ATD, reflecting its inflexible structure (Figure 2.9). The largest CCSs of the open-chain isomer $[M+3H]^{3+}$ and unknot macrocycle $[M+3H]^{3+}$ were 395 ± 3.5 $Å^2$ and 368 ± 5.3 $Å^2$, respectively. The largest CCS observed for the trefoil knot $[M+2H]^{2+}$ ion was 292 ± 1 $Å^2$. Therefore for the highest charge state observed for each species, the open-chain isomer has a larger molecular cross-section than the unknot macrocycle which, in turn, has a much larger cross-sectional area than the trefoil knot. Furthermore, the open-chain isomer has the broadest CCS distribution, followed by the unknot macrocycle, with the trefoil knot having the narrowest range. These results indicate that the trefoil knot has a much more compact and inflexible structure than the unknot macrocycle, which is more compact and less flexible than the precursor strand. Thus chemically identical isomers can be distinguished and characterised by IM-MS [3].

2.3.3. Multiple ATD Peak Fitting

For ATDs with multiple peaks which are not baseline resolved (Figure 2.12), values for average scan numbers (and hence average arrival times) of separate species can be obtained by fitting the ATDs with a Gaussian function in Origin 8.0 (OriginLab, Northampton, MA, USA). The Gaussian function used is stated in Equation 2.3, where offset $y_0 = 0$, A is the area, x_c the centre of the fitted peak and w the peak width:

$$y = y_0 + \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-2\frac{(x-x_c)^2}{w^2}} \qquad \textit{Equation 2.3}$$

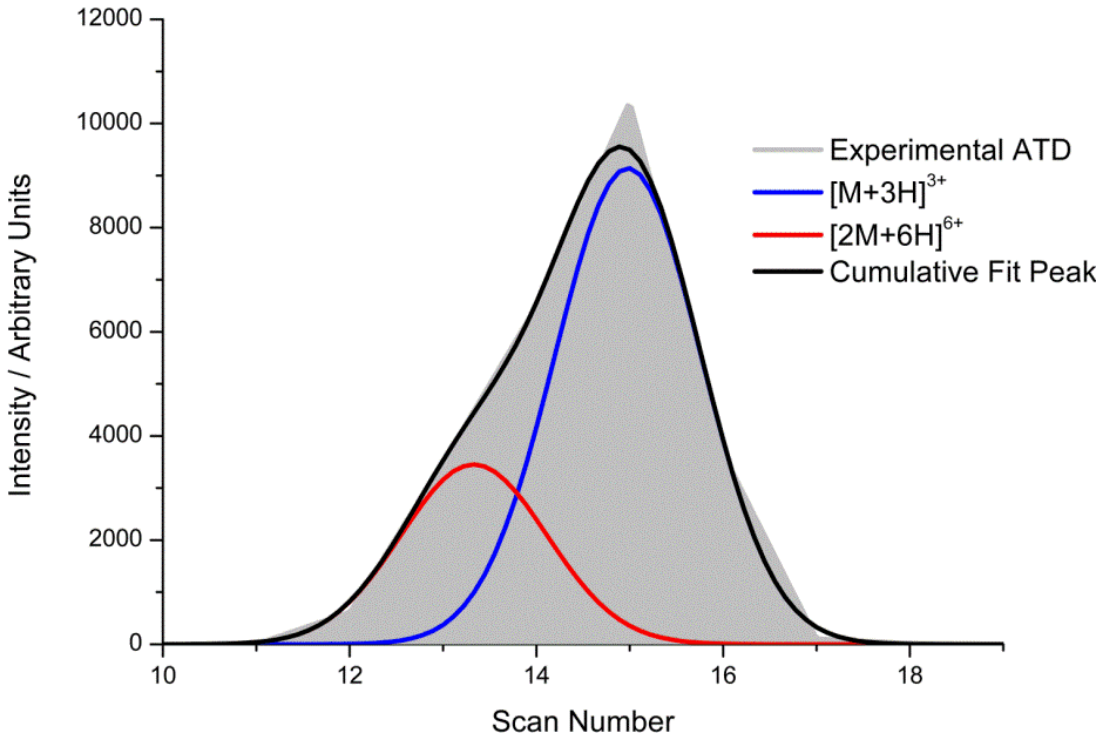


Figure 2.12 | Experimental ATD of insulin m/z peak 1912 at a DV of 50 V (grey shading). Two species are present; $[2M+6H]^{6+}$ and $[M+3H]^{3+}$. Gaussian fits (red and blue lines respectively) allow the centre of each species ATD to be calculated. The cumulative peak fit is the sum of the Gaussian fits and closely approximates the experimental ATD data.

The centre of the fitted peak x_c can be converted into the average arrival time for each species at this drift voltage. When x_c values have been obtained for all drift voltages a plot of *P/V* versus arrival time is made, as in Figure 2.10. Data processed in this way tend to have a lower R^2 value for the best fit line, so the >0.9995 criterion is relaxed to >0.995. As previously, the gradient is then used to calculate the rotationally averaged CCS of the molecular species. This method is comparable to the method used elsewhere in the literature where the ATD is fitted using the equation for the flux of ions exiting the drift tube [4].

2.4. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

2.4.1. Instrument

High resolution MS was performed on a 9.4 Tesla Apex Qe Fourier Transform Ion Cyclotron Resonance mass spectrometer (FT-ICR MS) (Bruker Daltonik GmbH). Differential pumping is employed for the transition of ions from atmospheric pressure to the ~2 x 10^{-11} mBar vacuum present in the analyser cell. The source region comprises a capillary (an in-house developed heated brass capillary to improve desolvation), skimmers and a hexapole. Ion optics transfer analyte ions from the hexapoles to the ICR cell.

Calibration of between 6 and 10 ppm accuracy was achieved using Bruker tune mix and data processed using DataAnalysis 4.0 (Bruker Daltonik GmbH). Simulated isotopic distributions were created from theoretical empirical formulas using the Simulate Isotopic Pattern function of DataAnalysis 4.0 (Bruker Daltonik GmbH). FT-ICR MS provides ^{13}C isotopic distributions which confirm the identification of the peaks as isobaric aggromers or as conformers of the same species.

2.4.2. Nano-Electrospray Ionisation Source Attachment

To ensure that FT-ICR MS source conditions emulated as closely as possible those of the QTOF and MoQTOF instruments utilised in other parts of the experiment a QTOF n-ESI source attachment was clamped in front of the entrance aperture (Figure 2.12). An external voltage was applied using a power pack to produce spray. The geometry and distance between the capillary and the entrance aperture was similar to that of a traditional Z Spray™ source.

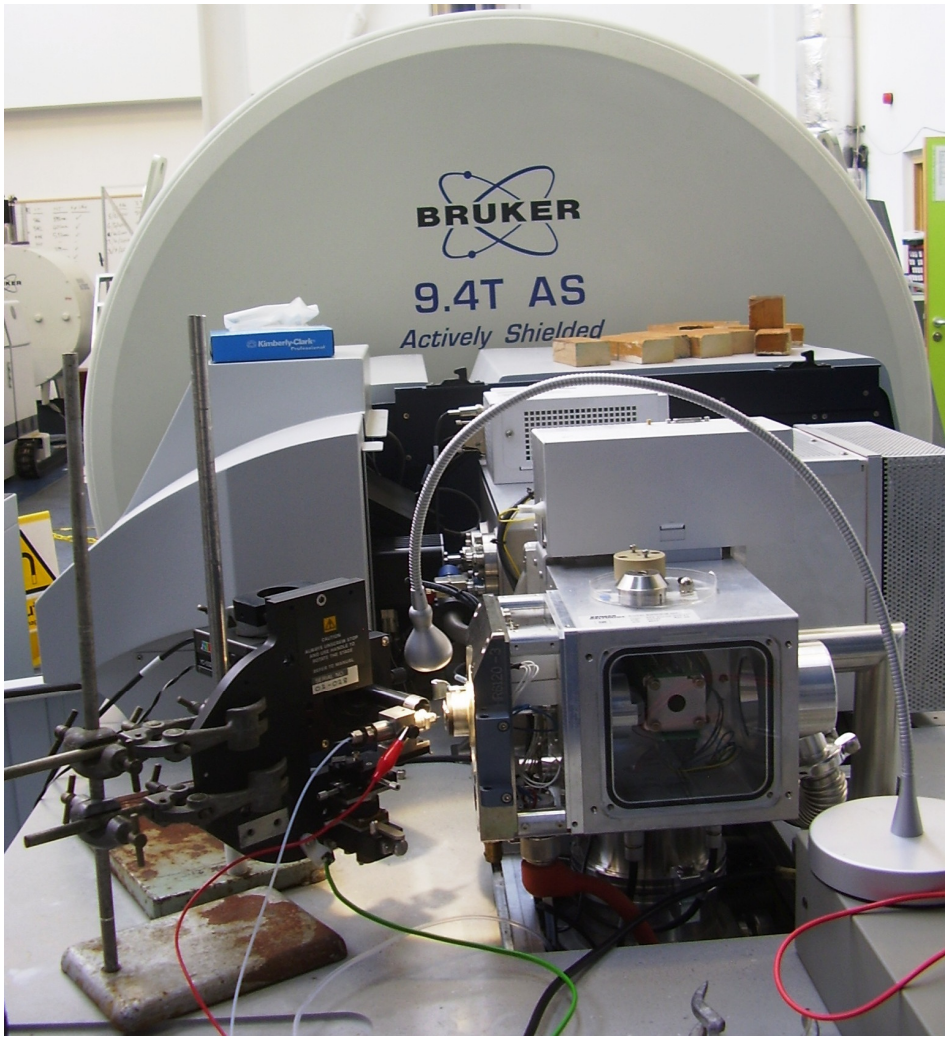


Figure 2.13 | Photo of the n-ESI source clamped in front of the FT-ICR MS entrance aperture.

2.5. References

1 McCullough, B. J. *et al.* Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal. Chem.* **80**, 6336-6344 (2008).

2 Macmillan, D. *et al.* Synthesis of Cyclic Peptides through an Intramolecular Amide Bond Rearrangement. *ChemBioChem* **12**, 2133-2136 (2011).

3 Barran, P. E. *et al.* Active-Metal Template Synthesis of a Molecular Trefoil Knot. *Angew. Chem. Int. Edit.* **50**, 12280-12284 (2011).

4 Bleiholder, C., Dupuis, N. F., Wyttenbach, T. & Bowers, M. T. Ion mobility mass spectrometry reveals a conformational conversion from random assembly to beta-sheet in amyloid fibril formation. *Nat. Chem.* **3**, 172-177 (2011).

3

Transthyretin (105-115)

Small fragments of amyloidogenic proteins can have fibrillogenic characteristics in vitro [1]. These amyloidogenic peptides are easier to synthesise and less computationally demanding to simulate, qualities which make them ideal for use as model systems to investigate amyloidogenesis. The evolution of oligomeric species formed by an aggregating solution of transthyretin (105-115) is examined using MS and IM-MS. Experimental results are compared with MD simulations to further elucidate oligomeric structures.

3.1. Introduction

3.1.1. Transthyretin

Transthyretin (TTR) is a 55 kDa homotetrameric protein [2], present in plasma and cerebrospinal fluid, which functions as a carrier for vitamin A (retinol) and thyroid hormone (thyroxine) to bodily tissues. Each 127 residue TTR monomer consists of eight β-strands, connected by loops, which are labelled conventionally from A to H, with orthogonal strands CBEF and DAGH separated by 10 Å [3]. Dimers are formed through β-strand interactions, resulting in hydrogen bonding between strands F and F' and H and H' of each monomer (Figure 3.1). Tetramer assembly is mediated by hydrophobic and hydrophilic interactions of the AB loop and H' strands (Figure 3.2). Higher aggregation requires interactions across the A strands, which have evolved to be aggregation resistant.

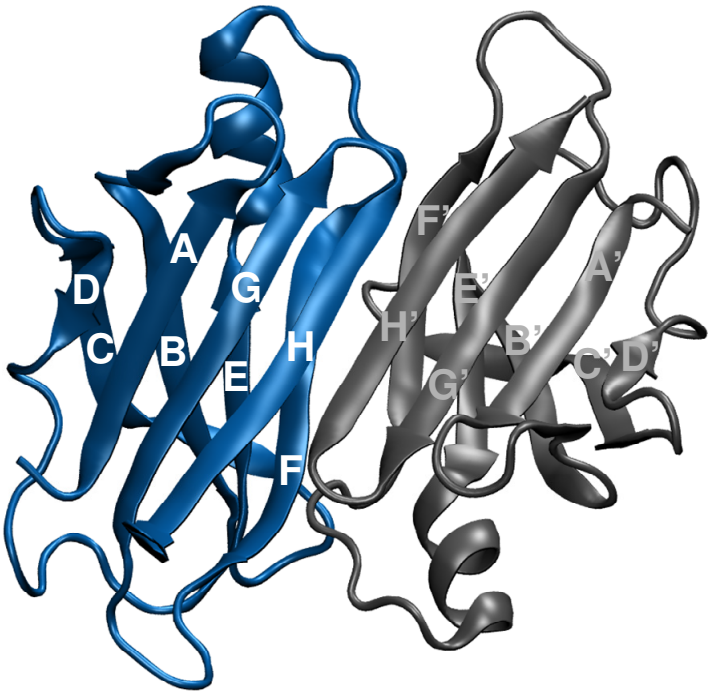


Figure 3.1 | TTR dimer with labelled β-strands. Representation rendered orthographically using VMD from pdb file 2PAB [3].

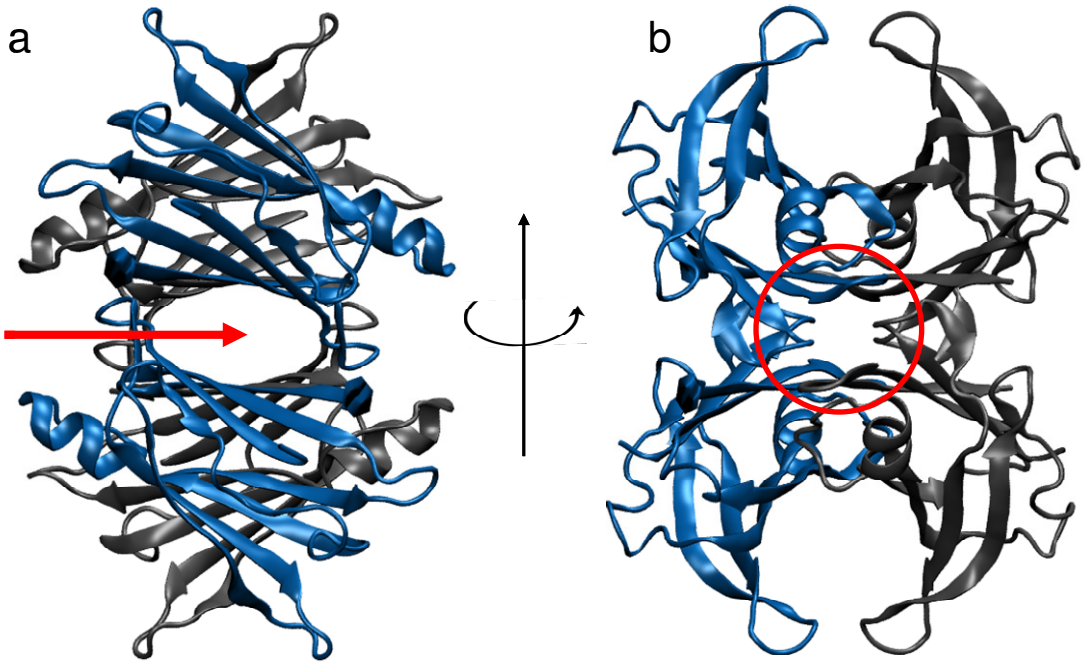


Figure 3.2 | Location of TTR tetramer binding sites: (a) red arrow denotes the position of the thyroxine binding channel (b) 90° rotation about the vertical axis, red circle shows the location of the retinol binding pocket. Representations of pdb file 2PAB [3] rendered orthographically using VMD.

Diseases associated with TTR aggregation are known as "Transthyretin amyloidoses". These include Familial Amyloid Polyneuropathy (FAP), Familial Amyloid Cardiopathy (FAC), Senile Systemic Amyloidosis (SSA) and Central Nervous SSA. Over 25 % of people over the age of 80 are estimated to be affected by SSA [4], which results from the aggregation of wild type (WT) TTR, fibrils of which are deposited in cardiac tissue [5]. Conversely FAP, one of the most common inherited amyloid diseases, is predominantly caused by a single point mutation in TTR, resulting in fibril deposition in the peripheral nervous system. There are over 80 known mutations of TTR which increase propensity to aggregation [6]; however crystal structures obtained for some of these variant TTR complexes show no significant structural differences from native TTR [7].

Fibrils present in FAP patient amyloid plaques have been studied by EM and X-ray diffraction. Initial studies found that fibrils consist of four protofilaments of between

40 and 50 Å in diameter aligned in parallel around a hollow centre [8]. The protofilaments themselves contain four β-sheets which twist together in a rope-like fashion parallel to the axis of the protofilament. Conflicting work observed *ex vivo* fibril diameters of between 80 and 120 Å in diameter, composed of four protofilaments of 20 Å diameters, which contained two β-sheets [9]. NMR coupled with HDX allowed Olofsson *et al.* [10] to investigate the fibril structure of a disease related TTR variant. The degree of residue protection from bulk solvent was ascertained, leading to the postulation of a fibril core of six β-strands A-B-E-F-G-H, with strands C and D exposed to solvent and A and B available for external associations. This is critical as one of the functions of the CD region is to protect the hydrophobic surfaces of the A B strands and prevent aggregation. Thus C and D strands, and any mutations therein leading to conformational variability, have a dramatic impact on the aggregation propensity [11,12]. The postulate that the mobility of the C D section is vital for protofibril assembly is gaining substantial experimental backing. However protofibril formation may only proceed after a conformational change, such as a slipped edge interaction [12].

A universal characteristic of the TTR fibrillation process is the initial destabilisation of the native homotetrameric structure [13]. TTR tetramer dissociation has been observed under non-denaturing conditions [14]. Modification of a QTOF mass spectrometer to allow the transmission of large macromolecular assemblies enabled Sobott *et al.* [15] to preserve the TTR homotetrameric structure, complexed with thyroxine and retinol, into the gas-phase. The role of retinol, in conjunction with the retinol binding protein, in stabilising the TTR homotetramer was elucidated by Hyung *et al.* [16] and is of particular import as physiological low levels of retinol binding protein have been recorded in FAP patients [17]. MS and HDX have been utilised to study tetramer and monomer populations in WT and variant TTR by Nettleton *et al*., who discovered a correlation between increasing amyloidogenicity and tetrameric instability in the gas-phase [18]. The presence of thyroxine in solution stabilised the tetramer, as did the presence of water in the central channel. Amyloidogenic variant TTR monomers were perceived to partially unfold their β-sheet structure, suggesting that monomer unfolding is a key step in the fibrillation

process. The stability of TTR has been studied also by molecular dynamics simulations [19] which supported the conclusions of Nettleton *et al.* [18] that the tetramer was stabilised by binding with thyroxine and additionally showed that solution conditions need to be neutral to favour the formation of tetramers.

The link between the instability of TTR tetramer mutants and the aggressiveness of their associated diseases was studied by Babbes *et al.* [20], judging tetramer stability by the ease with which urea denatured the structure. Tetramers were observed to first dissociate into monomers which then partially unfolded. A correlation was observed between the amyloidogenicity of the mutant form and the instability of the tetramer, in agreement with Nettleton *et al.* [18]. Relative to the wild type tetramer, the most aggressive variants were found to have destabilised quaternary and tertiary structures, whilst those of intermediate severity possessed either an unstable quaternary and stable tertiary structure, or vice versa. These studies and others have sparked interest from the pharmaceutical industry in finding ligands which could specifically bind to tetrameric TTR and stabilise the native state. Such compounds could be potential therapeutics for transthyretin amyloidoses [21].

The gas-phase dissociation of the tetramer was further probed by MS by Sobott *et al.* who report the tetrameric form dissociating, by means of collisions with buffer gas atoms, into highly charged monomeric and less charged trimeric ions [22]. Asymmetric dissociation was also observed for octameric ions. It was suggested that the tetramer fragments to form monomers which then associate into trimers. Collisional activation of the TTR tetramer followed by IM-MS analysis showed an expansion of the CCS prior to dissociation [23]. Simulations suggested the cause of this was the almost complete unfolding of one monomer subunit initially, followed by progressive unfolding of the remaining components at increasing activation energies. This provides evidence that partially unfolded complexes can remain intact on the millisecond experimental timescale and give insights into the mechanism of tetramer dissociation.

Multiple studies have reported that first steps in fibrillogenesis are the break up of the TTR homotetramer followed by the formation of a partially folded monomer, which acts as an aggregation intermediate [13,24-27]. At low pH, conditions which, *in vitro*, promote TTR aggregation [24], monomers represent the most populated species [25]. TTR reassembles into its native structure by an unusual monomer-dimer-trimer-tetramer pathway [28]. Assuming that TTR aggregation does progress *via* a monomer intermediate, both the reassembly and aggregation processes must be reliant on the concentration of monomer. For reassembly, the monomer must be folded and for aggregation to proceed it must be partially unfolded; it is postulated that monomer conformation depends on the concentrations of aggregates and reassembly intermediates present [28]. If concentrations of aggregates are high, an increased proportion of monomers partially unfold, leading to increased aggregation rates. Therefore in individuals where the rate of aggregate clearance is slower (such as the elderly) increased rates of aggregation occur, explaining the late onset and degenerative nature of amyloid disease.

Although deposition of fibrillar material can inhibit cell function, studies have found that TTR amyloid fibrils and large aggregates over 100 kDa in size were non toxic to cells. Instead preamyloidogenic cytotoxic forms of TTR were discovered [29], with further studies identifying these as monomeric species and rapidly formed transient oligomers up to hexameric order. This stresses the importance of further research into prefibrillar intermediates [30].

3.1.2. Fibrillogenic Transthyretin Fragments

Investigations into the fibrillogenic characteristics of synthesised TTR fragment peptides discovered two sequence regions, TTR (10-20) and TTR (105-115), which were capable of forming amyloid fibrils *in vitro* [31]. Below is the primary sequence of a TTR monomer. Cyan highlighting represents the presence of β-strands, magenta the presence of turns and yellow the presence of an α helix.

1 3 6 12 18 19 22 29 35 37 39 41 48
GPTGTGESKCPLMVKVLDAVRGSPAINVAVHVFRKAADDTWEPFASGKTSE

53 55 57 60 61 63 66 73 75 81 87 9899 101
SGELHGLTTEEEFVEGIYKVEIDTKSYWKALGISPFHEHAEVVFTANDSGPR

104 112 115 123 127
RYTIAALLSPYSYSTTAVVTNPKE

TTR (10-20) corresponds to β-strand A of the TTR monomer (Figure 3.3) which, in tetrameric form, lines the thyroxine binding channel. Fibrillogenesis can be induced by acidic solution conditions and the resulting fibrils are measured by EM to be 14 nm in diameter. Addition of urea to disrupt the fibrils enabled the measurement of protofilament diameters of 3 to 4 nm [32]. Further disruption of protofilaments with urea and high temperatures was unsuccessful; however their dissociation into monomeric peptide was achieved using high concentrations of TFE. The inference drawn from this is that peptides forming protofilaments are strongly bound together by hydrophobic and electrostatic forces whereas the interactions between protofilaments to form mature fibrils are weaker and hence more vulnerable to disruption.

TTR (105-115), corresponding to β-strand G of the TTR monomer, is more amyloidogenic than TTR (10-20), as it does not require acidified solution conditions for fibrillogenesis to occur. A more detailed review of TTR (105-115) continues below.

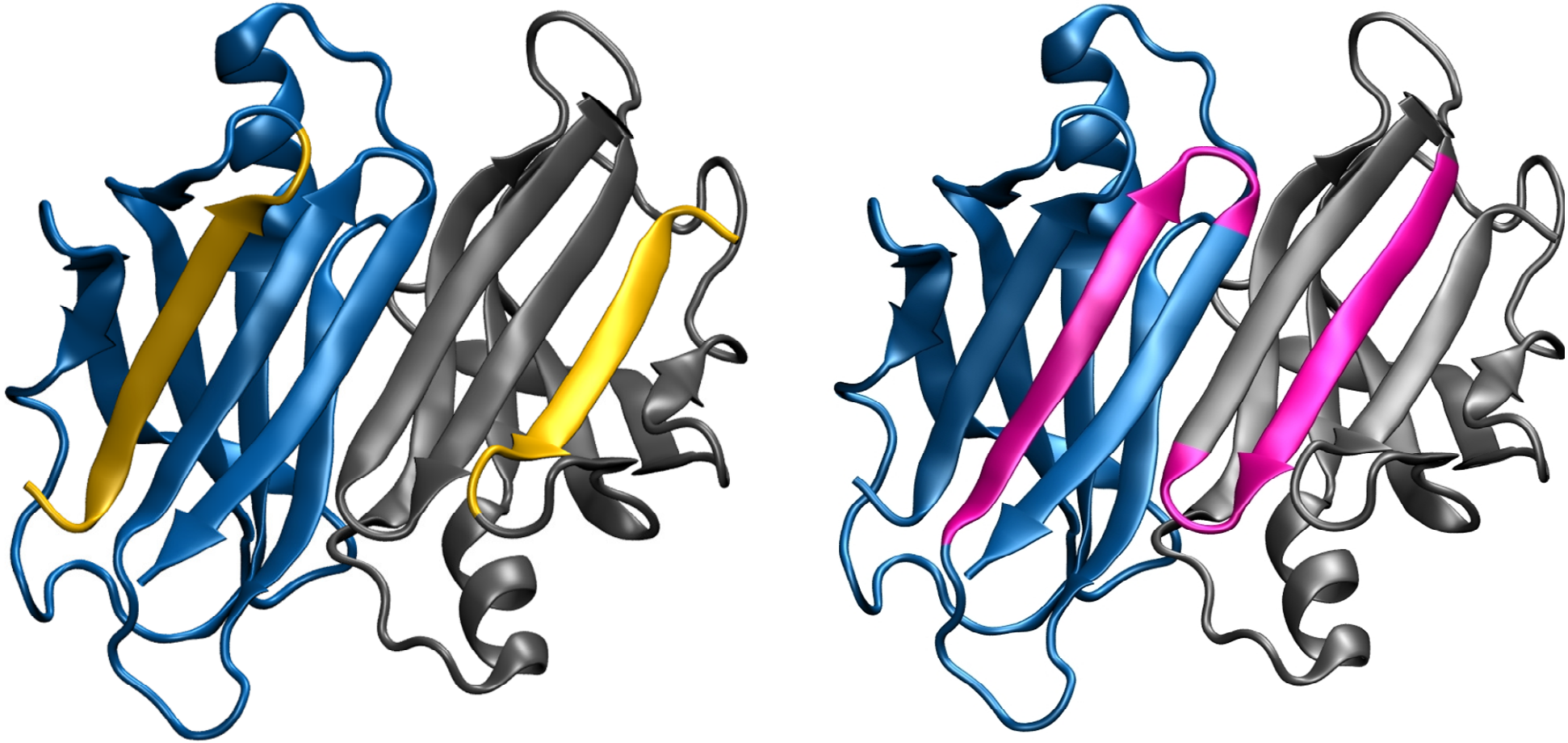


Figure 3.3 | TTR dimer showing the location of fibrillogenic sequence fragments in each monomer. Separate monomers are coloured blue and grey respectively. Residues (10 - 20) are highlighted in yellow and residues (105 – 115) in magenta. Representation rendered using VMD from pdb file 2PAB [3].

3.1.3. Transthyretin (105-115)

The amino acid sequence of residues from position 105 to position 115 of TTR is **Tyr-Thr-Ile-Ala-Ala-Leu-Leu-Ser-Pro-Tyr-Ser**. TTR (105-115) comprises β-strand G (Figure 3.3), which in the native tetramer is situated at the surface of the thyroxine channel (Figure 3.2). A high resolution structure of fibrillar TTR (105-115) has been discerned to atomic resolution by magic-angle spinning SSNMR spectroscopy (Figure 3.4) [33]. Residues in the fragment were isotopically labelled, enabling the calculation of the backbone torsion angles and distances. From these it was observed that the peptide was in extended β-strand conformation within the fibrils [34]. Differences between this and the similarly extended conformation adopted by residues (105-115) in the whole protein [3] can be accounted for by the greater impact of a proline residue in the shorter peptide sequence, causing a kink in the fragment polypeptide chain. More detailed studies determined a large number of NMR distances and dihedral angle restraints on the polypeptide chain including the conformation of side chains. All torsion angles were almost at their optimum values, allowing the peptide to have a molecular conformation of very low energy even in fibrillar form, a fact that would contribute greatly to high fibril stability and structural organisation. TTR (105-115) also possessed a degree of long range uniformity leading to the postulate that fibrils should be modelled as a polypeptide chain of crystalline form; a property supported by high fibril stability [33].

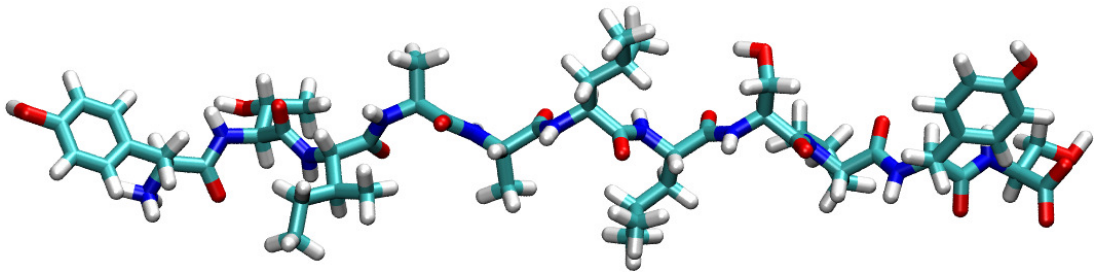


Figure 3.4 | Structure of fibrillar TTR (105-115). Orthographic representation rendered using VMD from pdb file 1RVS [33].

When TTR (105–115) fibrils are partially chemically denatured with urea and scanned by TEM, measurements of protofilament dimensions reveal an average diameter of 4.3 ± 1.6 nm [32]. NMR data calculates the length of the TTR fragment peptide between the tyrosine and serine carbon C′ atoms as 3.4 ± 0.2 nm, increasing by 0.4 nm if side chains are included [32]. The diameter of the protofilament is not much greater than the length of the peptide, suggesting that protofilament β-sheets are composed of a hydrogen-bonded array of TTR (105-115) monomers in an extended conformation. Results from X-ray fibre diffraction show a distance between β-strands of 0.47 Å and a β-sheet spacing of 0.8 to 1 nm [35]. Therefore four β-sheets in a "cross-β" configuration are postulated to compose each protofilament. Furthermore, as TEM provides an average diameter of 10.8 ± 1.2 nm for mature TTR (105–115) fibrils it suggests that a maximum of four protofilaments are wound together form the fibril structure [33].

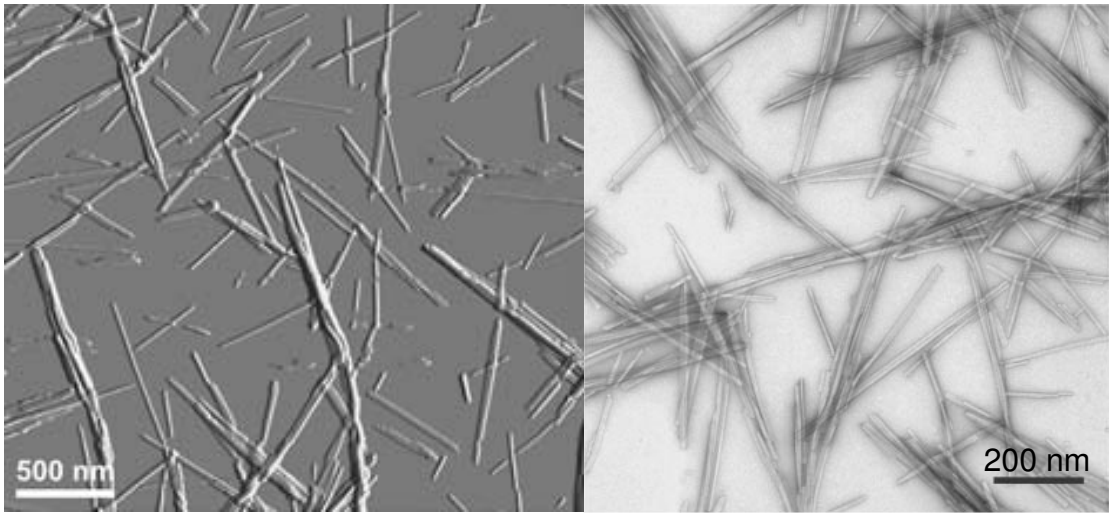


Figure 3.5 | TTR (105-115) fibrils imaged by AFM (left) [36] and EM (right) [34].

The morphology and mechanical strength of TTR (105-115) can be tested using AFM [36]. Although polymorphism is exhibited, the prevalent structure is straight fibrils of circa 1 μM in length and between 7 to 12 nm in diameter, with some displaying a surface striation every 90 nm perpendicular to the fibril axis [32]. The surface striation supports the assertion that mature fibrils are composed of at least two entwined protofilaments. Fibrils were unaffected by tapping mode AFM in air,

however in water fibrils were destroyed by AFM within 90 minutes. Contact mode AFM imaging in air allowed the strength of the fibrils to be precisely tested; at vertical forces below 40 nN fibrils were unaffected, increasing the force beyond this led to deformation, with all permanently distorted at applied forces of 100 nN. Finally the effect of solution conditions was investigated, with the fibrils retaining their structural stability from pH 3 to pH 8 in aqueous conditions and in sodium phosphate buffer concentrations between 50 mM and 1 M. AFM has also been used in conjunction with Fourier transform infrared spectroscopy (FTIR) to investigate fibril stability at high hydrostatic pressures [37]. Mature fibrils withstood hydrostatic pressures of up to 1.3 GPa, whilst the oligomers preceding fibril formation could be dissociated at 220 MPa. The implication is that early oligomers are only weakly bonded by predominantly electrostatic and hydrophobic interactions, whereas fibrils are robust as they have a large number of hydrogen bonds stabilising the structure from which bulk water is excluded.

Concerns that the observed cross-β structure of fibrils was a result of the dehydrated conditions used to observe previous samples were allayed by an X-ray scattering study into hydrated samples of fibrils [38]. Inter-strand and inter-sheet reflection properties were shared by both hydrated and dried samples over various concentrations. Furthermore the hydrated samples showed that strands were axially aligned to the fibril axis whilst the β-sheets were equatorially aligned, a configuration indicative of a cross-β structure. Interestingly this outcome implies that the β-sheets, when stacked together, must have several interfaces which are protected from bulk water, in agreement with Dirix *et al.* [37]. These results were crucial to allowing the presumption of a fibril cross-β structure in studies conducted with experimental techniques which require a sample solution, as opposed to dehydrated fibrils.

All-atom 5 μs MD simulations have been constructed to analyse the formation and structure of TTR (105-115) dimers, trimers and tetramers preceding fibrillogenesis [39]. Peptides self associated into β rich structures, with two different formation pathways for aggregates dominated by β structure and those dominated by α-helical structure. Anti-parallel conformations of β-strands were favoured,

supported by the propensity of the central residues (106-111) to form inter-peptide hydrogen bonds. The use of experimentally calculated distances and dihedral restraint angles [33] increased the peptide's aggregation propensity. Chemical shift data from experiment and simulated data from the generated structures were in agreement, illustrating that MD may aid in discriminating between potential structures for oligomers and fibrils. The interstrand re-arrangements in TTR (105-115) dimers were also studied by MD in implicit solvent [40]. The study found that it was probable that both parallel and antiparallel strand arrangements would be present although parallel aggregates were separated by energy barriers depending on their precise alignment. Interstrand rearrangements could take place in a variety of ways, including differing reptation movements.

The fibrillogenesis of TTR (105-115) has been observed using FRET probes [41,42]. These can detect fibril structural arrangements by virtue of changing FRET efficiency, complemented by direct EM observations of fibril formation. Parallel arrangements of TTR (105-115) within a β-sheet were postulated, with adjacent sheets arrayed in an antiparallel conformation. An intermediate state was observed which underwent a conformational compaction on production of mature fibrils. Cell adhesion and control sequences attached to C-terminus of TTR (105-115) peptides which undergo fibrillogenesis produce fibrils which are bioactive and interact specifically with cells [43]. The development of functionalised fibrils which could interact specifically with many cell varieties could have important medical and bio-mimetic uses.

In this chapter we extend current approaches by using MS and IM-MS to examine the change in the population of oligomers present in an aggregating solution of TTR (105-115). The distribution of oligomeric species at a zero time point is identified using n-ESI MS and quantified by IM-MS ATD intensity. Experimental CCSs are compared with those obtained by molecular modelling. Finally changes in the population of oligomers and their CCSs as a function of time are observed and analysed.

3.2. Methodology

3.2.1. Sample Preparation

Peptide TTR (105-115) was obtained as a lyophilised powder at >98 % purity from CSBio (USA). Unaggregating conditions were dilute (10 to 50 μM) solutions in 49.5:49.5:1 H_2O:ACN:HCOOH, and in buffer conditions with a higher percentage of ACN. To create an aggregating solution the peptide was dissolved at a concentration of 10 mg/ml in 90:10 H_2O:ACN adjusted to pH 2 with HCl. These samples were immediately flash frozen in liquid nitrogen and stored in a -28 $^{\circ}$C freezer. For time-course experiments a 10 mg/ml sample was defrosted and aliquots were taken at appropriate time points and diluted to a concentration of 80 μM in a solution of 49.5:49.5:1 H_2O:MeOH:HCOOH immediately prior to mass spectrometry.

3.2.2. Mass Spectrometry

Mass spectrometry was performed on a QTOF 2 mass spectrometer (Waters, Manchester, UK). Mass spectra were recorded for samples in positive mode utilising an n-ESI ionisation source. High resolution mass spectra provided ^{13}C isotopic distributions which confirmed the identification of peaks where several assignations might have been made due to the presence of oligomers with coincident *m/z* values. Source voltages and pressures were tuned to enable optimal aggregate transmission.

3.2.3. Ion Mobility Mass Spectrometry

IM-MS measurements were performed on an in-house modified QTOF 1 (Waters, Manchester, UK), the 'MoQTOF' [44]. More details of this instrument can be found in Chapter 2, Section 2.3.1. Samples were ionised using an n-ESI source in positive mode with an applied capillary voltage of ~1.8 kV and a cone voltage of 50 V.

Source pressure was increased to optimise transmission of oligomers, as other studies have reported [15]. Ions were accelerated through a potential difference of 35 V and injected into the drift cell. The drift cell was filled with helium gas at a pressure of between 3.0 and 3.5 Torr. The temperature and pressure of helium was recorded at the start and end of each mobility scan and the average taken. The voltage applied across the cell was varied from 60 V to 15 V and ATDs recorded at eight voltages. The resolution of the MoQTOF is not sufficient to give isotopic ^{13}C spacing. Therefore oligomeric species were assigned using potassium adduct spacing from the mass spectrum in addition to information gleaned from ATDs.

3.2.4. Simulation Strategies

All simulations were performed by Jason Kalapothakis [45,46]. MD simulations utilised the AMBER9 programs [47], employing ff99 force field parameters. Simulation temperature was regulated with the Berendsen velocity thermostat [48]; a 1 fs time-step was used in all simulations and the SHAKE algorithm [49] was used to prevent unrealistic bond and angle stretching. The energy minimisation process consisted of 50,000 iterations, first 100 steepest descent minimisation steps and the remainder employing the conjugate gradient algorithm. Minimisation was halted when the convergence criterion (0.5 cal/mol) was satisfied.

Analysis of the resulting molecular coordinates was performed with ptraj, included in the AMBER programs. CCSs were estimated with MOBCAL using the trajectory method (TM) [50].

3.3. Results and Discussions

3.3.1. Mass Spectrometry

3.3.1.1. Aggregation Inhibiting Conditions

Conditions *in vitro* which inhibit TTR (105-115) aggregation are low peptide concentrations and buffer conditions containing high percentages of organic solvents.

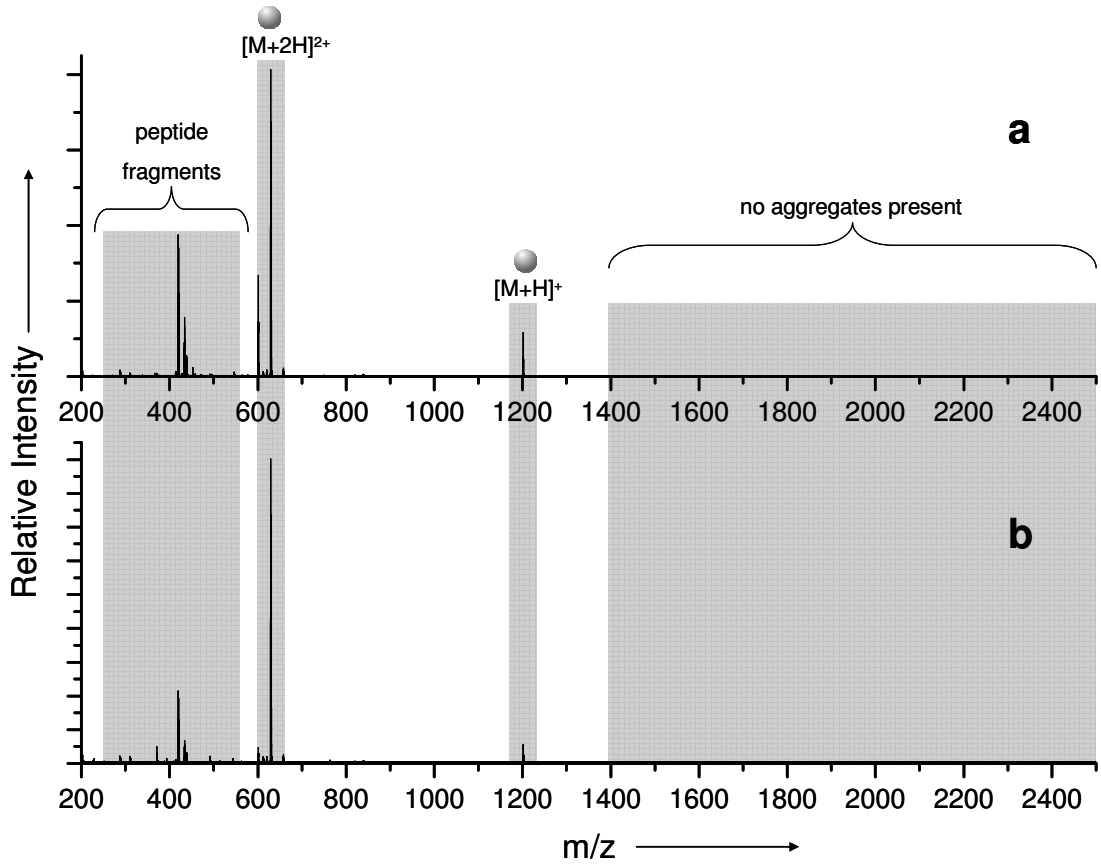


Figure 3.6 | n-ESI mass spectra of 40 µM TTR (105-115) in conditions which inhibit aggregation. ***a****. buffer conditions of 74.5:24.5:1 ACN:H_2O:HCOOH.* ***b****. buffer conditions of 49.5:49.5:1 ACN:H_2O:HCOOH.*

Under non-aggregating conditions the peptide is observed exclusively as a monomer, predominantly as a $[M+2H]^{2+}$ ion with a smaller proportion of $[M+H]^{+}$ ion (Figure 3.6).

When a sample of 10 mg/ml in 90:10 H_2O:ACN adjusted to pH 2 with HCl is defrosted and left to aggregate, then diluted to a concentration of 40 μM in a solution of 49.5:49.5:1 H_2O:MeOH:HCOOH prior to mass spectrometry, a wide variety of oligomers are observed, in opposition to Figure 3.6b. Therefore oligomer populations present in the more concentrated aggregating stock solution must persist under more dilute, non aggregating, conditions.

3.3.1.2. Aggregation Promoting Conditions

Samples of 10 mg/ml TTR (105-115) in solvent conditions of 90:10 H_2O:ACN adjusted to pH 2 with HCl were observed to aggregate without agitation at room temperature, inferred from an increase in solution viscosity with time. Prior to MS analysis samples were diluted to a concentration of 80 μM in a solution of 49.5:49.5:1 H_2O:MeOH:HCOOH. The oligomer populations observed in the diluted solution are relevant to those present in the stock solution, as exemplified by the persisting oligomer populations which are described above when an aggregating stock solution is diluted into non aggregating conditions.

Figure 3.7 shows a typical n-ESI mass spectrum obtained from a solution of the TTR (105–115) peptide immediately after dilution.

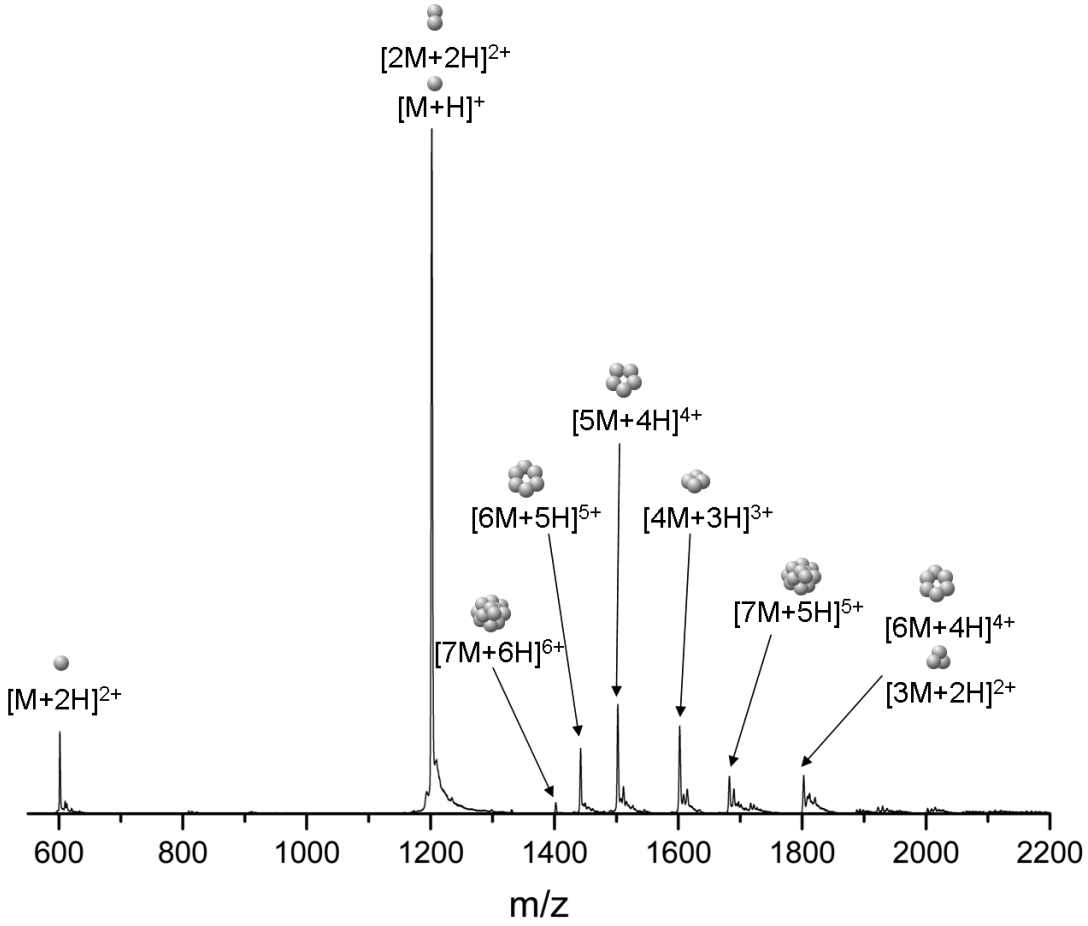


Figure 3.7 | n-ESI mass spectrum showing oligomers present and their associated charge states. Cartoons are for illustrative purposes only and do not correspond to validated structures.

Oligomers of the general form $[nM+zH]^{z+}$ are observed with $1 \leq n \leq 13$ (Figure 3.7 and Figure 3.8). With the notable exceptions of the work of Ashcroft *et al.* on β_2m [51] and that of Nettleton *et al.* on bovine insulin [52], our observation of a wide range of oligomers is atypical of amyloidogenic systems. The spectrum is dominated by a peak at *m/z* 1202.0, corresponding to coincident species $[M+H]^+$ and $[2M+2H]^{2+}$, with other high-abundant ions at *m/z* 601.5, 1502.4, 1601.8, and 1441.8 assigned to $[M+2H]^{2+}$, $[5M+4H]^{4+}$, $[4M+3H]^{3+}$ and $[6M+5H]^{5+}$ respectively. Each oligomer is found in relatively few protonation states, which is similar to the findings of Bowers *et al.* for Aβ [53]. With the exception of $[M+2H]^{2+}$, $[M+H]^+$ and $[2M+2H]^{2+}$, all of the observed species carry fewer charges than the number of peptides in the oligomer.

Large numbers of charges increase the CCS of an ion due to Coulombic repulsion [54,55] thus the absence of highly charged assemblies indicates that the observed ions either adopt compact geometries or are extended but possess a low number of ionisable sites. IM-MS analysis can determine which explanation is correct.

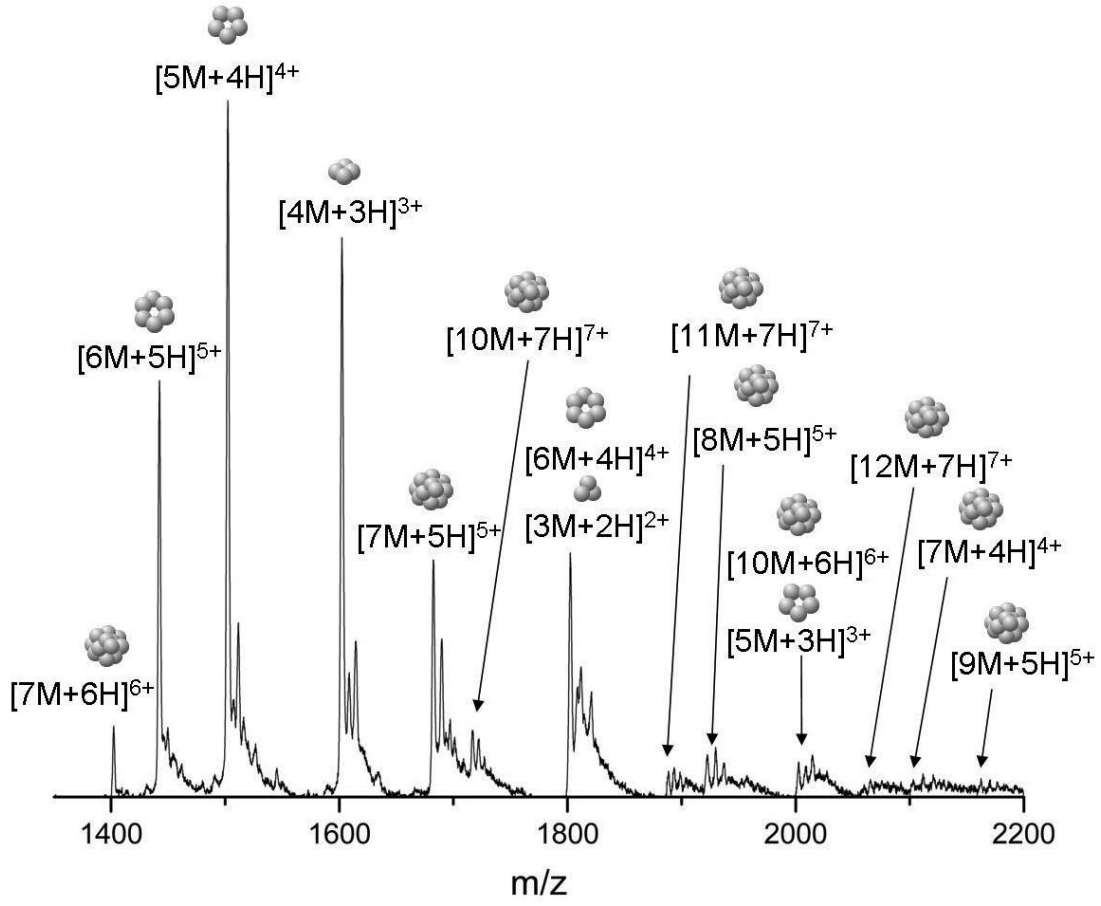


Figure 3.8 | n-ESI mass spectrum showing an enlargement of the high m/z region of the mass spectrum showing oligomers present and their associated charge states. Cartoons are for illustrative purposes only and do not correspond to validated structures.

3.3.2. Ion Mobility Mass Spectrometry

3.3.2.1. Zero Time Point Experiments

The summed intensity of each ion's ATD was recorded after the same number of acquisitions. This enables the consideration of the relative abundances of each oligomer.

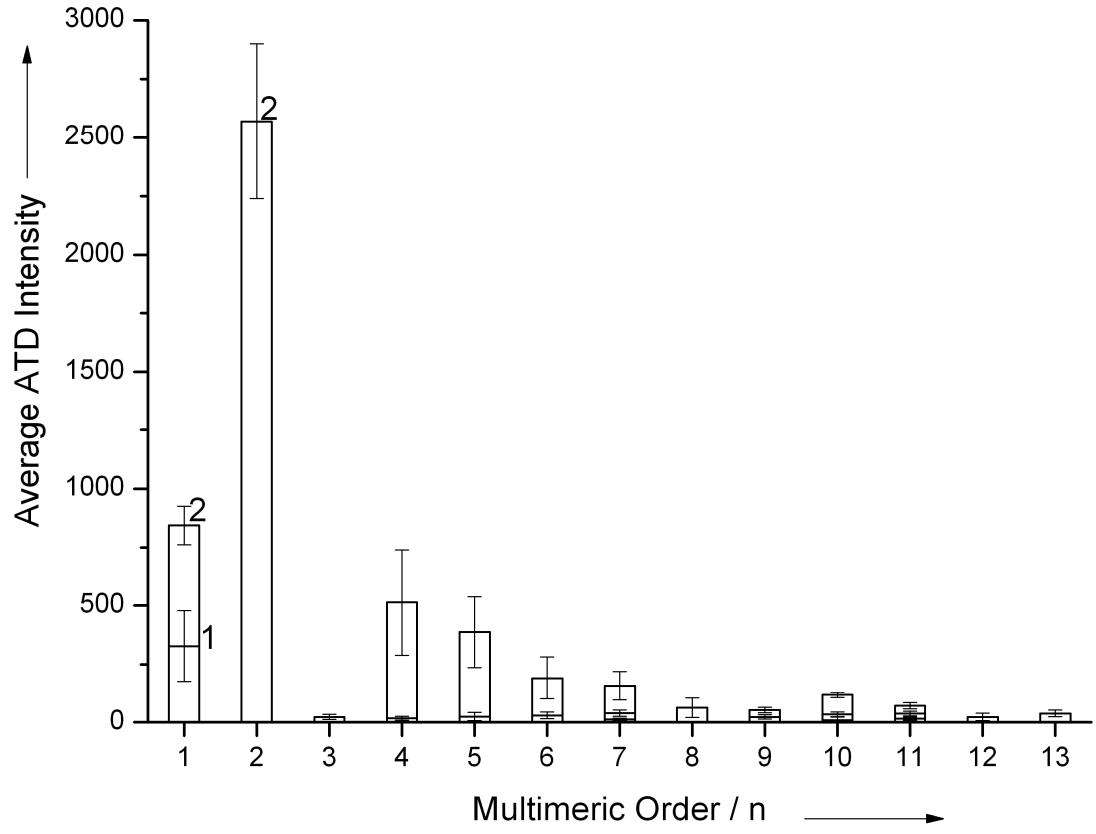


Figure 3.9 | Distribution of oligomers at t = 0. Each bar represents the average maximum intensity of the ATD of each charge state of each oligomer. All ATDs were recorded at a drift voltage of 50 V. Labels correspond to the oligomeric charge state, +z. The data are averaged over three repeats; error bars represent the standard deviation.

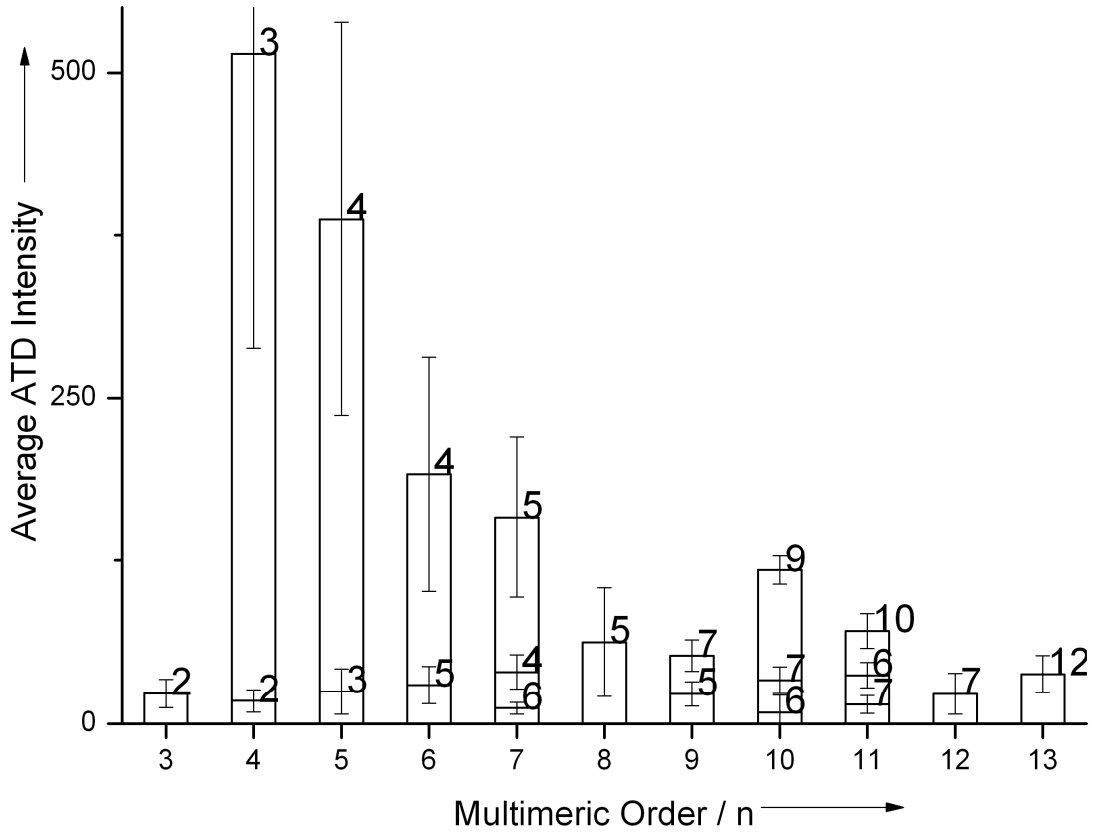


Figure 3.10 | Distribution of oligomers at t = 0, showing an expansion of the region for n > 2. Each bar represents the average maximum intensity of the ATD of each charge state of each oligomer. All ATDs were recorded at a drift voltage of 50 V. Labels correspond to the oligomeric charge state, +z. The data are averaged over three repeats; error bars represent the standard deviation.

When the data are viewed in this way, it is obvious that some species are dominant (Figure 3.9). The peak at 1202.0 corresponds mainly to $[2M+2H]^{2+}$, thus the dimer is the most abundant species under aggregating conditions. The tetramer and pentamer are also seen in high abundance, the former solely as the ion $[4M+3H]^{3+}$ and the latter principally as the ion $[5M+4H]^{4+}$. When more than one charge state is observed, one value of *z* typically dominates. A very weak signal from the trimer is visible, which points to an instability for this species relative to the dimer and tetramer, in support of simulation results [39]. Contributions from other oligomers

decay exponentially, with the exception of the $[10M+zH]^{z+}$, which appears as a stable 'magic-number' oligomer [56]. The presence of 'magic number' oligomers with increased stability strengthens the argument that the oligomers observed are present in solution, and not artefacts of the n-ESI process.

Figure 3.10 shows that the higher-order aggregates are observed in unequal charge states: for example $[10M+zH]^{z+}$ is seen with z = 6, 7, and 9 but not 8; $[9M+zH]^{z+}$ is observed with z = 5 and 7 but not 6; and $[11M+zH]^{z+}$ is present as z = 6, 7, and 10 but not 8 and 9. Charge states that are not observed may be too weak to be detected and for some *m/z* values are coincident with very abundant species of lower oligomeric order; nonetheless they appear to have a lower stability that may be due to the adoption of favourable conformations by only some oligomers. We speculate that this is due to antiparallel or parallel arrangements of polypeptide chains that favour only certain charge and aggregation states.

A simulation study by Paci *et al.* found the tetramers of TTR (105-115) were unstable [39], suggesting that the critical nucleus for amyloidogenesis is of greater oligomeric order than a tetramer. Given the stable nature and unusual prevalence of the decameric oligomers in these experiments, we postulate that this could be a candidate species for the critical nucleus. We can speculate that a decameric species could possibly also be one of the cytotoxic oligomers which characterise many aggregation processes. CCSs were determined for all of the species present. Results from this analysis are shown in Figure 3.11 and tabulated in Appendix Three.

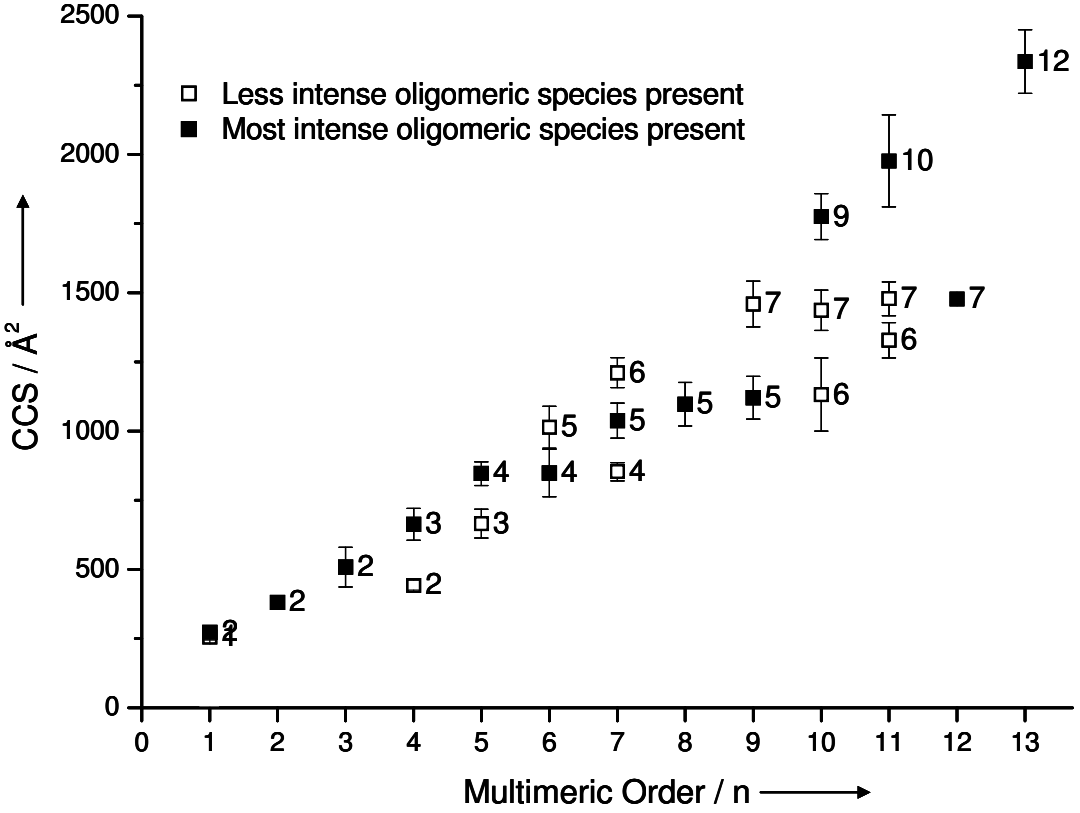


Figure 3.11 | Graph showing the CCSs of all species observed at a zero time point. Labels are the associated positive charge of each oligomer. Error bars represent the standard deviation of five repeated measurements.

As expected, an increase in CCS with multimeric order is observed. In addition, where an oligomer is present in a range of charge states, CCS increases with charge. However variation from a monotonic rate of increase in CCS with oligomeric order can allow assumptions concerning structural information to be made. The monomeric peptide is seen as both a singly and a doubly charged ion, with very similar CCSs of 256.1 Å² and 272.7 Å² respectively. The dimer however is seen in a single charge state of z = 2 and has a CCS of 380.2 Å², significantly larger than the monomer species. The weak $[3M+2H]^{2+}$ ion has a CCS of 508.5 Å², larger than the dimer and twice the size of the monomer. Oligomer CCSs increase in a linear fashion for the highest charge state for each oligomer, which is to be expected, but what is more

interesting is the lack of increase in CCS as a function of n when $n > 3$. For example, the CCSs for $n = 4$ and 5 are very similar when $z = 3$ (Figure 3.11), and the CCSs of the hexamer, heptamer, octamer and nonamer are within 10 % when $z = 5$ despite an increase in mass of >60 %.

Thus packing to form a dense aggregate competes with Coulombic repulsion to determine the final CCS of an oligomer. Within any oligomer, the amino group at the N-terminus of each peptide is available for protonation. It is also possible that the proline residue can be protonated, although this is less probable as it has a lower basicity than the N-terminus. The protonation of both the amino group at the N-terminus and the proline is likely to account for the observed $[M+2H]^{2+}$ species. If this protonated group is packed in the interior of the aggregate and hence inaccessible, the oligomer will carry a low net charge relative to the number of peptides; if it is surface-exposed, ionization will drive charge–charge repulsion, Coulombic destabilisation, and expansion of the aggregate. Our observation that all of the species carry fewer charges than the number of peptides in the oligomer, as well as the comparatively small increase in CCS with a large increase in mass, suggests dense packing. Moreover, given the sequence of TTR (105–115), it is likely that this packing is governed by hydrophobic interactions; as these are weakened in the gas-phase, the species detected must have been preserved intact from solution.

3.3.2.2. Time Course Experiments

Oligomer populations were monitored during the lag phase of TTR (105-115) aggregation. Figure 3.12 and Figure 3.13 show mass spectra taken at intervals over an eight hour period, which is within the lag phase of TTR (105-115) self-assembly [33]. For the first four hours there is a rise in the observable oligomer population followed by a rapid decline in most of the aggregates, with the exception of the monomer, dimer (Figure 3.12) and tetramer species (Figure 3.13), which persist over the period of study.

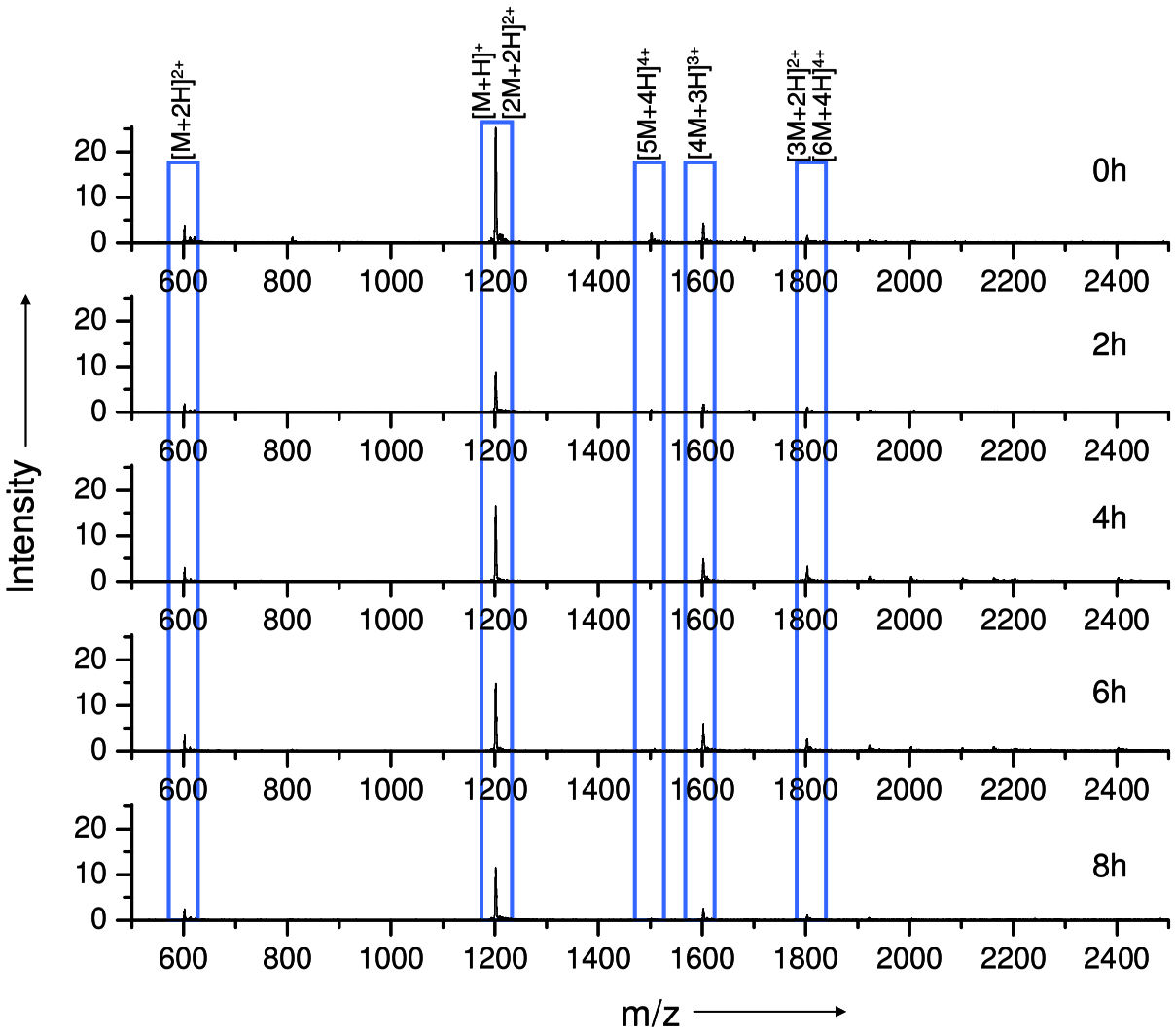


Figure 3.12 | MoQTOF time course spectra. Created by scanning over the same number of scans at a drift voltage of 50 V. Raw data are processed by first subtracting the background: parameter values of Polynomial Order 5, Below Curve 20 % and Tolerance 0.010. The data are smoothed six times over 15 channels using the Savitzky Golay method. Spectra shown are representative of all data gathered.

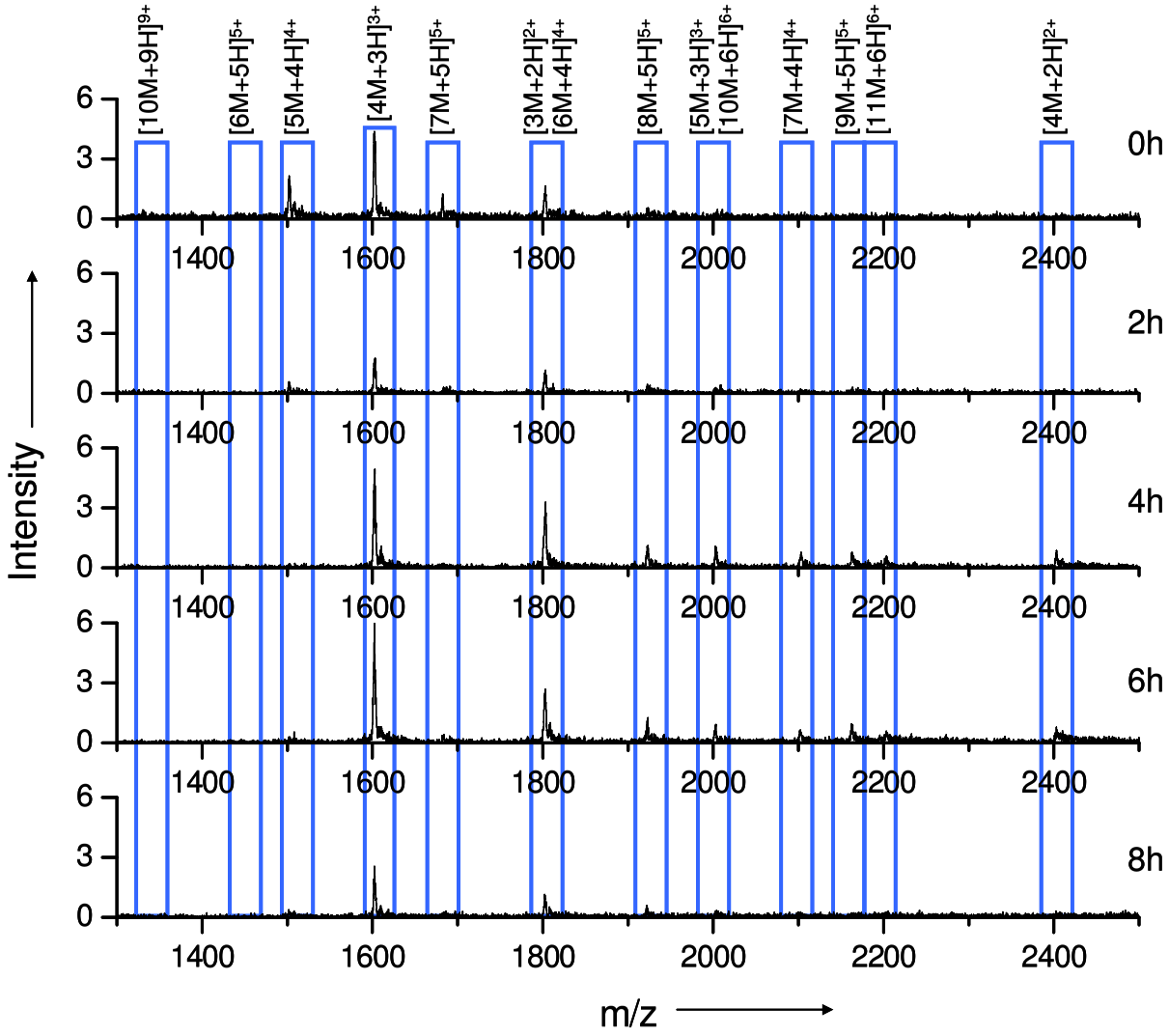


Figure 3.13 | MoQTOF time course spectra, an enlargement of the higher m/z region. Created by scanning over the same number of scans at a drift voltage of 50 V. Raw data are processed by first subtracting the background: parameter values of Polynomial Order 5, Below Curve 20 % and Tolerance 0.010. The data are smoothed six times over 15 channels using the Savitzky Golay method. Spectra shown are representative of all data gathered.

The *m/z* ratios of some oligomeric species are coincident. However, these species may be separated in mobility space by their differing arrival times, revealing that the ratio of intensities of the two populations varies with time. The ATDs of coincident species at zero, two, four, six and eight hours at a drift voltage of 50 V are shown below, averaged over three repeats.

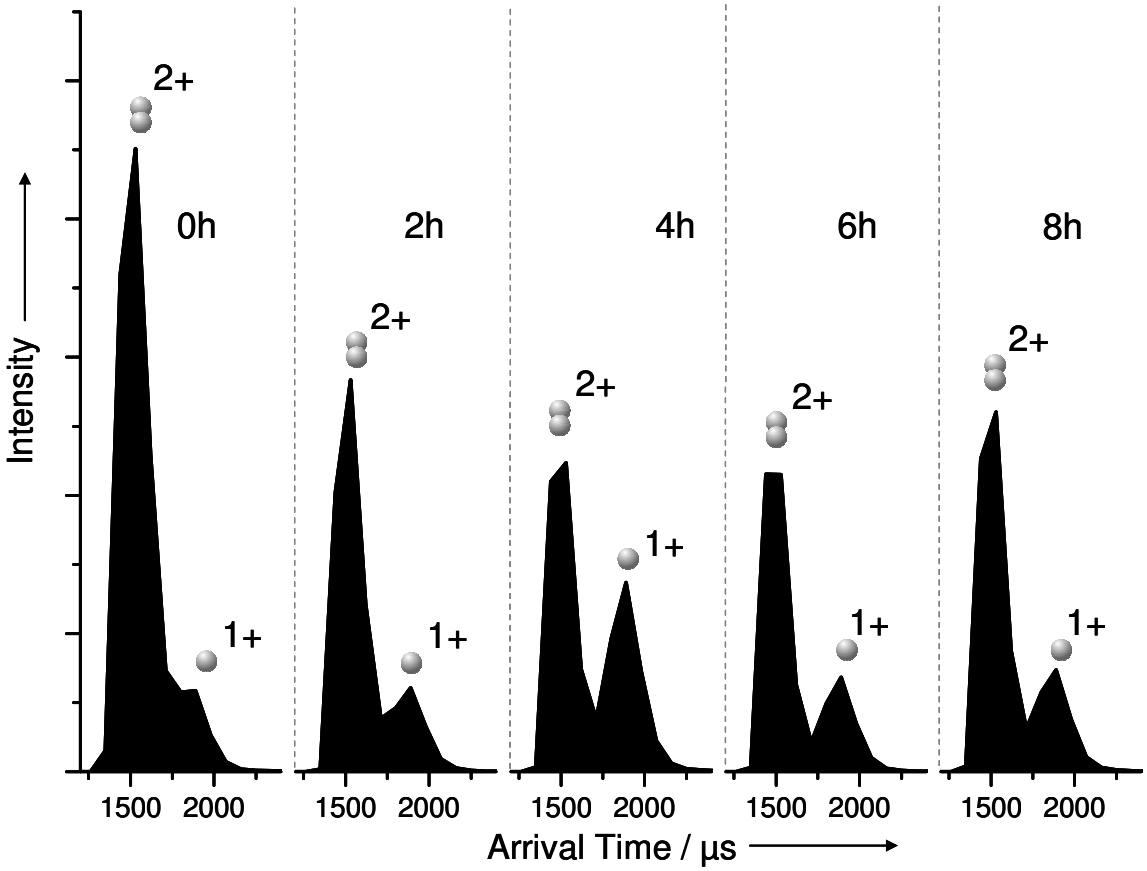


Figure 3.14 | The arrival time distributions, at a drift voltage of 50 V, of $[M+H]^+$ and $[2M+2H]^{2+}$ at zero, two, four, six and eight hours.

When the dominant peak at 1202.0 is viewed in ATD space, the proportion attributable to dimer and monomer is seen to vary with time (Figure 3.14). Initially there is far more $[2M+2H]^{2+}$ present than $[M+H]^+$; however, at four hours relative enrichment of the monomer is observed. At both the six hour and eight hour time

points, the proportion of dimer is seen to rise again. The abundance of monomer at 4 hours most probably originates from the break-up of larger oligomers owing to destabilisation either in solution or upon desolvation.

These data, combined with those shown in Figure 3.12 and Figure 3.13, show a particular and significant persistence of dimer and tetramer whilst the other more transient oligomers contribute either to aggregation or off-pathway processes. The presence of the tetramer in an on-pathway process would be consistent with previous data, suggesting that the minimal unit of the mature fibril is assembled from four β-sheets in a "cross-β" configuration [33].

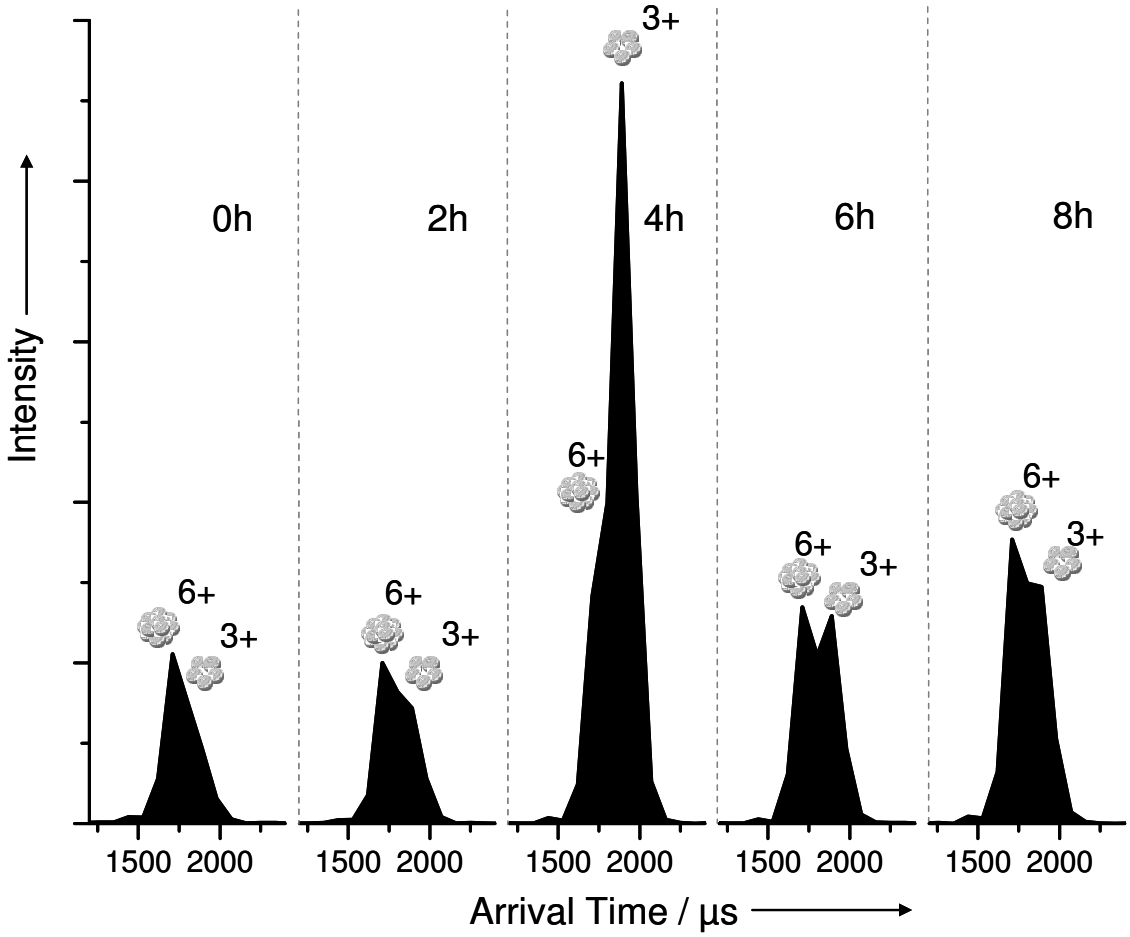


Figure 3.15 | The arrival time distributions, at a drift voltage of 50 V, of $[5M+3H]^{3+}$ and $[10M+6H]^{6+}$ at zero, two, four, six and eight hours.

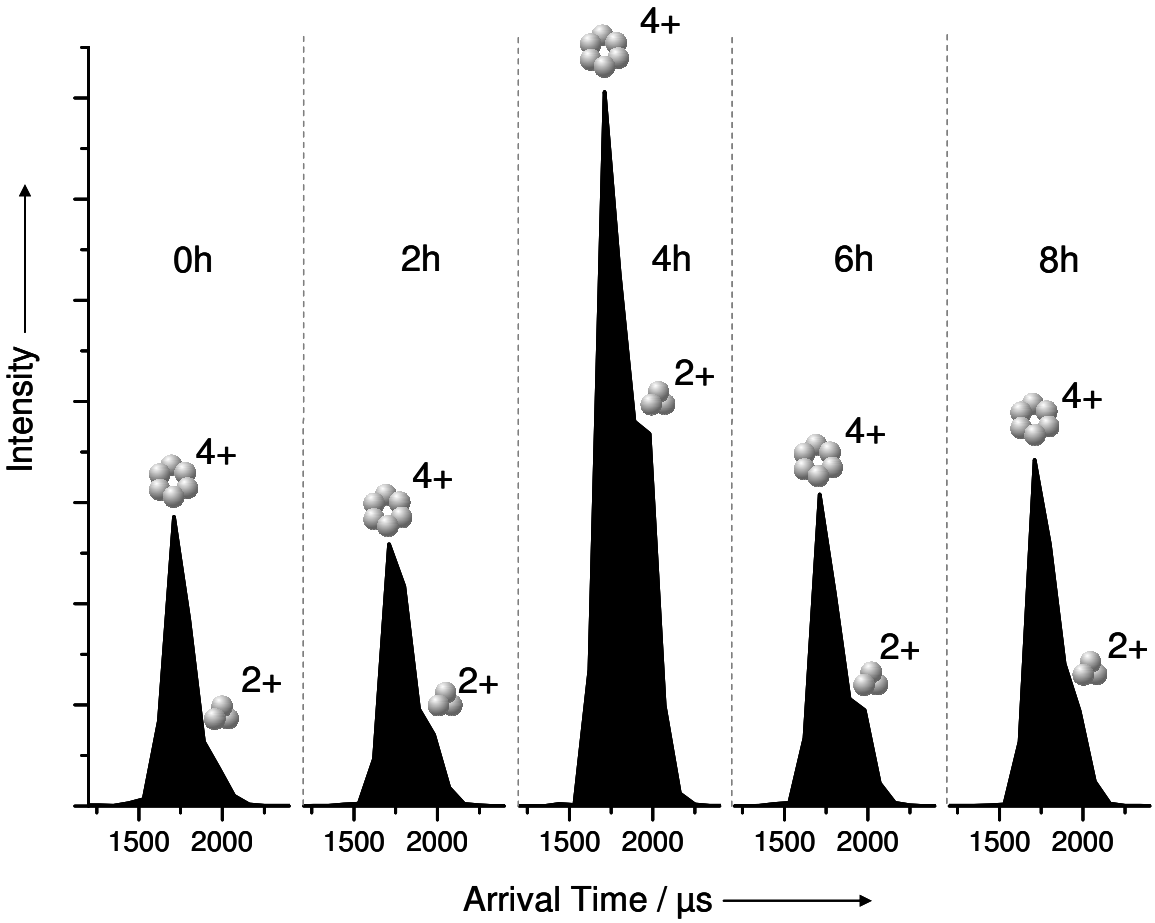


Figure 3.16 | The arrival time distributions, at a drift voltage of 50 V, of $[3M+2H]^{2+}$ and $[6M+4H]^{4+}$ species at zero, two, four, six and eight hours.

At the four hour time point a peak is seen in the intensity of the lower multimeric order species in relation to the corresponding higher order *m/z* coincident species (Figure 3.15 and Figure 3.16). As a relative enrichment of monomeric species is observed at this time point (Figure 3.14) it seems likely that the increase in the population of smaller oligomeric species corresponds to a break-up of unstable higher order aggregates at this time point due to destabilisation either in solution or upon desolvation.

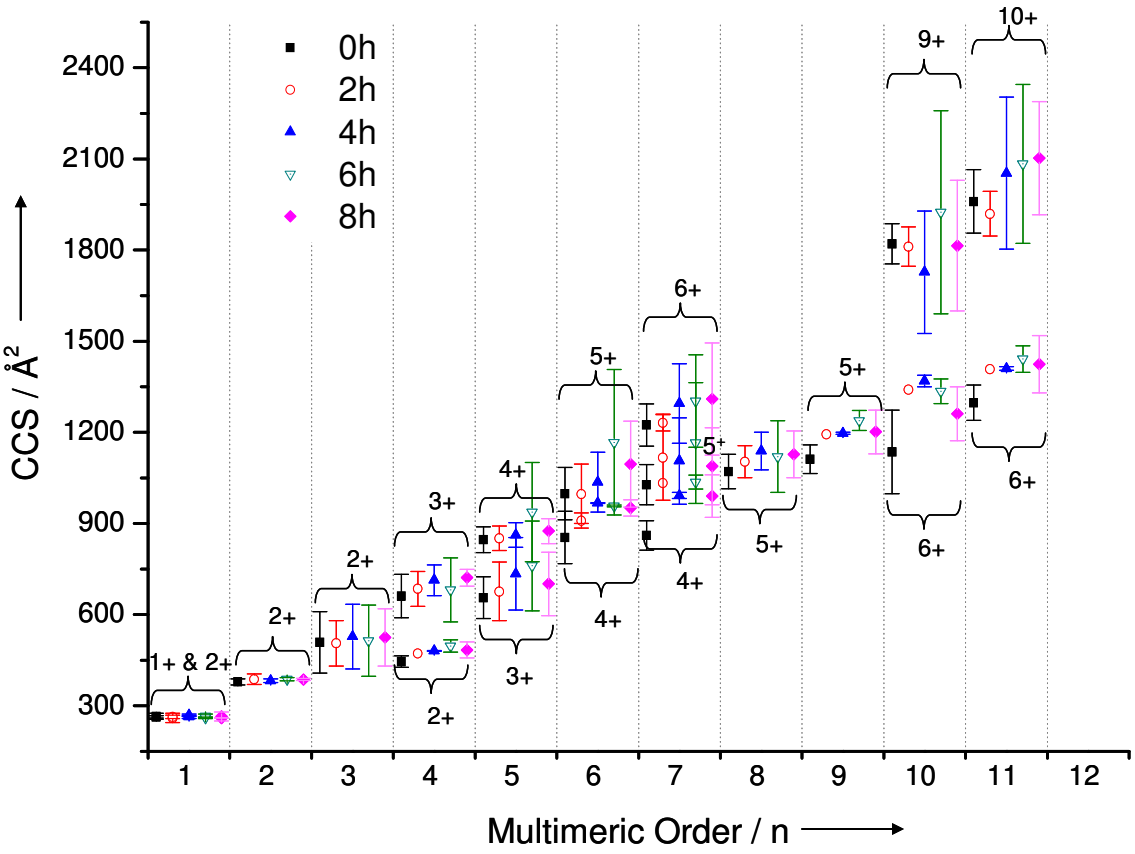


Figure 3.17 | CCS evolution of each oligomer as a function of time.

Plotting the CCS as a function of time demonstrates a systematic increase in CCS for the majority of oligomers with $n \geq 4$ (Figure 3.17), whilst the CCSs remain constant for the smaller oligomers. This observation is consistent with dynamic rearrangement of the larger $n \geq 4$ oligomers and the adoption of increasingly extended, possibly β-strand-like, conformations.

3.3.3. Simulations

Experimental data from IM-MS measurements provide a coarse grained shape which can be compared with atomistically resolved structures derived either *via* computational routes or from NMR or crystal structures. Given atomic co-ordinates, it is possible to calculate a CCS that can be compared to an experimental measurement [50,55,57].

Probable candidate geometries for TTR (105–115) oligomers were modelled using two approaches [46]. In the first oligomers of monomeric to hexameric order, with antiparallel β-sheet structural components were 'cut out' of solid state NMR data [32,33]. These were subjected to molecular mechanics at 300 K for 5 ns. The second method utilised a simulated annealing (SA) approach to generate 300 gas-phase structures for oligomers, with the lowest energy structures considered as likely candidates. With the exception of the monomer, which contained a cationic N-terminus and a neutral C-terminus giving rise to $[M+H]^+$, all simulated oligomers were of the form $[nM+(n-1)H]^{(n-1)+}$, achieved by addition of an anionic C-terminal carboxyl in one peptide. This neutrally charged zwitterionic peptide was placed in the centre of oligomers larger than dimeric order, allowing the anionic C-terminus to be bordered by positively charged N-termini of neighbouring peptides, thus increasing the stability of the array. Oligomers from monomeric to hexameric order were simulated. Point charges for neutral C-terminal residues were derived using the RESP procedure [58].

Model isolated, ionised, antiparallel β-sheets were energy-minimised (Section 3.2.4). After energy minimisation heating was conducted up to 350 K in 50 K steps over 800,000 MD iterations (0.8 ns) and subsequently 'relaxed' in the gas-phase over 5,000,000 MD iterations (5 ns). Coordinates before and after MD were used for further analysis.

The same sets of initial coordinates were also submitted to several cycles of SA in order to generate collapsed gas-phase structures. Peptides were first heated abruptly to a high temperature (800 K for the monomeric peptide and 600 K for multimeric assemblies) for 30,000 MD iterations (30 ps). Subsequently the system is cooled linearly to 0 K over 20,000 MD iterations (20 ps) after which it is energy-minimised. Peptide assemblies were prevented from drifting apart at high temperatures by application of a flat-bottomed harmonic restraint between the centres of mass of individual molecules with a cut off distance of 40 Å. This procedure was repeated numerous times (typically 500) and the coordinates resulting at the end of each SA cycle saved.

Both approaches resulted in collapsed globular structures for the monomer and dimer, with no β-sheet structure present. Following 5 ns of MD the trimer retained its β-sheet structure, present in the NMR data; however this was not the case after SA, which engendered a collapse. This trend was repeated for tetrameric and pentameric oligomers. Candidate structures are depicted in supplementary information of Cole *et al.* [46]. Theoretical CCSs were estimated with MOBCAL using the TM for structures produced by both methods.

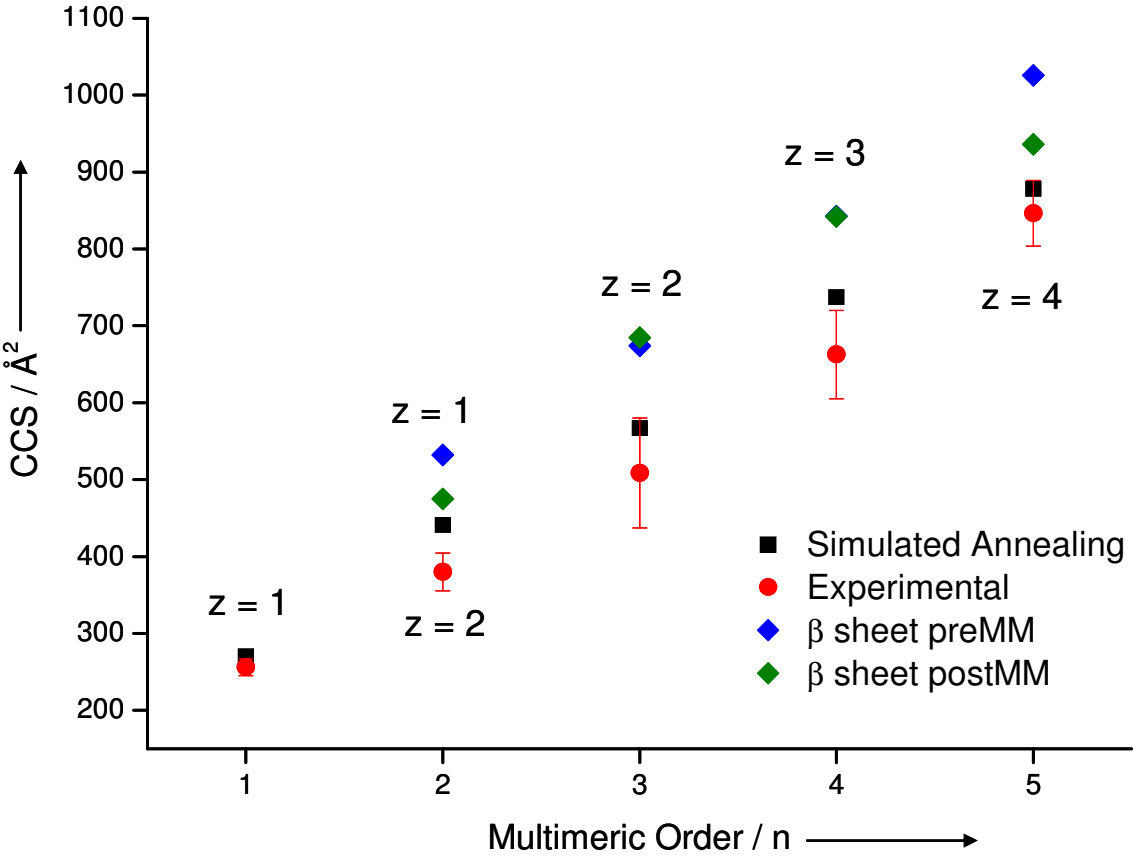


Figure 3.18 | Graphical comparison between experimental CCSs and those generated by SA, in addition to β-sheet structures pre and post molecular mechanics. Labels refer to the positive charge of each oligomer.

Evidently from Figure 3.18 oligomer structures generated by SA are more representative of those observed by experiment. A direct comparison of SA CCSs and experimental CCSs is shown in Table 3.1.

	Collision Cross Section / Å²		
Species	**Theoretical**	**Experimental**	**% Difference**
$[M+H]^+$	270.0	256.1	4.3
$[2M+H]^+$	441.1	380.2*	6.4
$[3M+2H]^{2+}$	567.0	508.5	14.1
$[4M+3H]^{3+}$	737.0	662.7	8.7
$[5M+4H]^{4+}$	878.0	846.5	5.0

*Table 3.1 | Comparision between the experimental and theoretically calculated CCSs. *experimental data not available for $[2M+H]^+$, CCS is that of $[2M+2H]^{2+}$.*

The overall agreement between the two sets of data is good. The disparity between the theoretical $[2M+H]^+$ and the experimental $[2M+2H]^{2+}$ CCSs can be accounted for by the $[2M+2H]^{2+}$ species having a more collapsed structure due to increased attractive electrostatic interactions. Experimentally observed CCSs are consistently smaller than the calculated structures and increase in CCS with oligomeric order at a rate lower than that of the SA structures. Lower experimental CCSs could be due to injection energy effects or the presence of unresolved higher order oligomeric ions under the ATDs. Another contribution to this could be the presence of a protonated proline, a form not considered in these calculations, and more extensive simulations could produce structures in better agreement with experimental results.

The comparison of experimentally derived CCSs with simulations demonstrates that the oligomers observed are compact globular structures with no extended β-sheet. Experimental CCSs suggest that the solution phase structures are even more densely packed than those observed by simulation, a supposition supported by the narrow charge state distributions exhibited by each oligomer. Analysis of structures [45] produced by SA suggests that TTR (105-115) monomers re-arrange, substituting intramolecular interactions with intermolecular non-covalent interactions.

3.4. Summary

IM-MS provides a detailed insight into the molecular associations that ultimately lead to amyloid-like fibril formation. The amyloidogenic TTR (105–115) peptide forms a large variety of aggregates during the early stages of fibrillogenesis. Densely packed oligomers are observed, which are dominated by hydrophobic interactions, and dynamic remodelling of these oligomers occurs during the lag phase prior to fibril assembly. It is not possible to assess which of the many species are responsible for on pathway aggregation leading to amyloid fibril assembly, however of particular note is our observation of a stable tetrameric species, the oligomeric unit which is consistent with protofilament composition. Also of note is our detection of a stable, compact, and unusually prevalent decamer. It is commonly accepted that oligomers formed early in the aggregation process are toxic to a wide variety of cell types and that disease may result from the pathogenic effects of these “toxic oligomers”; we speculate that the transiently stable, compact, and prevalent decamer is a candidate for such a species.

3.5. References

1 Halverson, K., Fraser, P. E., Kirschner, D. A. & Lansbury, P. T. Molecular Determinants of Amyloid Deposition in Alzheimers Disease - Conformational Studies of Synthetic Beta-Protein Fragments. *Biochemistry* **29**, 2639-2644 (1990).

2 Blake, C. C. F. *et al.* X-Ray Study of Subunit Structure of Prealbumin. *J. Mol. Biol.* **61**, 217 (1971).

3 Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. Structure of Pre-Albumin - Secondary, Tertiary and Quaternary Interactions Determined by Fourier Refinement at 1.8A. *J. Mol. Biol.* **121**, 339-356 (1978).

4 Westermark, P., Sletten, K., Johansson, B. & Cornwell, G. G. Fibril in Senile Systemic Amyloidosis is Derived from Normal Transthyretin. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2843-2845 (1990).

5 Gustavsson, A. *et al.* Amyloid Fibril Composition and Transthyretin Gene Structure in Senile Systemic Amyloidosis. *Lab. Invest.* **73**, 703-708 (1995).

6 Connors, L. H., Richardson, A. M., Theberge, R. & Costello, C. E. Tabulation of transthyretin (TTR) variants as of 1/1/2000. *Amyloid-J. Protein Fold. Disord.* **7**, 54-69 (2000).

7 Hornberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A. E. A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J. Mol. Biol.* **302**, 649-669 (2000).

8 Blake, C. C. F. *et al.* A Molecular Model of an Amyloid Fibril in *Symposium on the Nature and Origin of Amyloid Fibrils.* 6-21 (John Wiley & Sons Ltd, Chichester, UK, 2007)

9 Inouye, H. *et al.* Analysis of x-ray diffraction patterns from amyloid of biopsied vitreous humor and kidney of transthyretin (TTR) Met30 familial amyloidotic polyneuropathy (FAP) patients: axially arrayed TTR monomers constitute the protofilament. *Amyloid* **5**, 163-174 (1998).

10 Olofsson, A., Ippel, J. H., Wijmenga, S. S., Lundgren, E. & Ohman, A. Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J. Biol. Chem.* **279**, 5699-5707 (2004).

11 Eneqvist, T. & Sauer-Eriksson, A. E. Structural distribution of mutations associated with familial amyloidotic polyneuropathy in human transthyretin. *Amyloid* **8**, 149-168 (2001).

12 Laidman, J., Forse, G. J. & Yeates, T. O. Conformational change and assembly through edge beta strands in transthyretin and other amyloid proteins. *Accounts Chem. Res.* **39**, 576-583 (2006).

13 Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J. M. & Brito, R. M. M. Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J. Biol. Chem.* **276**, 27207-27213 (2001).

14 Quintas, A., Saraiva, M. J. M. & Brito, R. M. M. The tetrameric protein transthyretin dissociates to a non-native monomer in solution - A novel model for amyloidogenesis. *J. Biol. Chem.* **274**, 32943-32949 (1999).

15 Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402-1407 (2002).

16 Hyung, S. J., Deroo, S. & Robinson, C. V. Retinol and Retinol-Binding Protein Stabilize Transthyretin via Formation of Retinol Transport Complex. *ACS Chem. Biol.* **5**, 1137-1146 (2010).

17 Shoji, S. & Nakagawa, S. Serum Prealbumin and Retinol-Binding Protein Concentrations in Japanese-Type Familial Amyloid Polyneuropathy. *Eur. Neurol.* **28**, 191-193 (1988).

18 Nettleton, E. J. *et al.* Protein subunit interactions and structural integrity of amyloidogenic transthyretins: Evidence from electrospray mass spectrometry. *J. Mol. Biol.* **281**, 553-564 (1998).

19 Sorensen, J., Hamelberg, D., Schiott, B. & McCammon, J. A. Comparative MD analysis of the stability of transthyretin providing insight into the fibrillation mechanism. *Biopolymers* **86**, 73-82 (2007).

20 Babbes, A. R. H., Powers, E. T. & Kelly, J. W. Quantification of the thermodynamically linked quaternary and tertiary structural stabilities of transthyretin and its disease-associated variants: The relationship between stability and amyloidosis. *Biochemistry* **47**, 6969-6984 (2008).

21 Sacchettini, J. C. & Kelly, J. W. Therapeutic strategies for human amyloid diseases. *Nat. Rev. Drug Discov.* **1**, 267-275 (2002).

22 Sobott, F., McCammon, M. G. & Robinson, C. V. Gas-phase dissociation pathways of a tetrameric protein complex. *Int. J. Mass Spectrom.* **230**, 193-200 (2003).

23 Ruotolo, B. T. *et al.* Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Edit.* **46**, 8001-8004 (2007).

24 Colon, W. & Kelly, J. W. Partial Denaturation of Transthyretins is Sufficient for Amyloid Fibril Formation In Vitro. *Biochemistry* **31**, 8654-8660 (1992).

25 Lai, Z. H., Colon, W. & Kelly, J. W. The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **35**, 6470-6482 (1996).

26 Liu, K., Cho, H. S., Lashuel, H. A., Kelly, J. W., Wemmer, D. E. A glimpse of a possible amyloidogenic intermediate of transthyretin. *Nat. Struct.Biol.* **7**, 754-757 (2000).

27 Lashuel, H. A., Lai, Z. H. & Kelly, J. W. Characterization of the transthyretin acid denaturation pathways by analytical ultracentrifugation: Implications for wild-type, V30M, and L55P amyloid fibril formation. *Biochemistry* **37**, 17851-17864 (1998).

28 Wiseman, R. L., Powers, E. T. & Kelly, J. W. Partitioning conformational intermediates between competing refolding and aggregation pathways: Insights into transthyretin amyloid disease. *Biochemistry* **44**, 16612-16623 (2005).

29 Sousa, M. M., Cardoso, I., Fernandes, R., Guimaraes, A. & Saraiva, M. J. Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy - Evidence for toxicity of nonfibrillar aggregates. *Am. J. Pathol.* **159**, 1993-2000 (2001).

30 Reixach, N., Deechongkit, S., Jiang, X., Kelly, J. W. & Buxbaum, J. N. Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2817-2822 (2004).

31 Gustavsson, A., Engstrom, U. & Westermark, P. Normal Transthyretin and Synthetic Transthyretin Fragments form Amyloid-Like Fibrils In Vitro. *Biochem. Biophys. Res. Commun.* **175**, 1159-1164 (1991).

32 MacPhee, C. E. & Dobson, C. M. Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *J. Mol. Biol.* **297**, 1203-1215 (2000).

33 Jaroniec, C. P. *et al.* High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 711-716 (2004).

34 Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M. & Griffin, R. G. Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16748-16753 (2002).

35 Jarvis, J. A., Craik, D. J. & Wilce, M. C. J. X-Ray-Diffracion Studies of Fibrils formed from Peptide-Fragments of Transthyretin. *Biochem. Biophys. Res. Commun.* **192**, 991-998 (1993).

36 Mesquida, P., Riener, C. K., MacPhee, C. E. & McKendry, R. A. Morphology and mechanical stability of amyloid-like peptide fibrils. *J. Mater. Sci. Mater. Med.* **18**, 1325-1331 (2007).

37 Dirix, C., Meersman, F., MacPhee, C. E., Dobson, C. M. & Heremans, K. High hydrostatic pressure dissociates early aggregates of TTR105-115, but not the mature amyloid fibrils. *J. Mol. Biol*. **347**, 903-909 (2005).

38 Squires, A. M. *et al.* X-ray scattering study of the effect of hydration on the cross-β structure of amyloid fibrils. *J. Am. Chem. Soc.* **128**, 11738-11739 (2006).

39 Paci, E., Gsponer, J., Salvatella, X. & Vendruscolo, M. Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. *J. Mol. Biol.* **340**, 555-569 (2004).

40 Li, D. W., Han, L. & Huo, S. Structural and pathway complexity of beta-strand reorganization within aggregates of human transthyretin(105-115) peptide. *J. Phys. Chem. B* **111**, 5425-5433 (2007).

41 Deng, W., Cao, A. & Lai, L. Detecting the inter-peptide arrangement and maturation process of transthyretin (105-115) amyloid fibril using a FRET pair with short Forster distance. *Biochem. Biophys. Res. Commun.* **362**, 689-694 (2007).

42 Deng, W., Cao, A. & Lai, L. Distinguishing the cross-beta spine arrangements in amyloid fibrils using FRET analysis. *Protein Sci.* **17**, 1102-1105 (2008).

43 Gras, S. L. *et al.* Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials* **29**, 1553-1562 (2008).

44 McCullough, B. J. *et al.* Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal. Chem.* **80**, 6336-6344 (2008).

45 Kalapothakis, J. *Investigations of Peptide Structural Stability in vacuo* PhD thesis, University of Edinburgh, (2009).

46 Cole, H. L., Kalapothakis, J. M. D., Bennett, G., Barran, P. E. & MacPhee, C. E. Characterizing Early Aggregates Formed by an Amyloidogenic Peptide by Mass Spectrometry. *Angew. Chem. Int. Edit.* **49**, 9448-9451 (2010).

47 Pearlman, D. A. *et al.* Amber, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **91**, 1-41 (1995).

48 Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-Dynamics with Coupling to an External Bath *J. Chem. Phys.* **81**, 3684-3690 (1984).

49 Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical-Integration of Caresian Equations of Motion of a System with Constraints - Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).

50 Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**, 16082-16086 (1996).

51 Smith, A. M., Jahn, T. R., Ashcroft, A. E. & Radford, S. E. Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* **364**, 9-19 (2006).

52 Nettleton, E. J. *et al.* Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophys. J.* **79**, 1053-1065 (2000).

53 Bernstein, S. L. *et al.* Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature Chem.* **1**, 326-331 (2009).

54 Clemmer, D. E. & Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass Spectrom.* **32**, 577-592 (1997).

55 Jarrold, M. F. Peptides and proteins in the vapor phase. *Annu. Rev. Phys. Chem.* **51**, 179-207 (2000).

56 Yamada, Y. & Castleman, A. W. The Magic Numbers of Metal and Metal Alloy Clusters. *J. Chem. Phys.* **97**, 4543-4548 (1992).

57 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86-91 (1996).

58 Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method using Charge Restraints for Deriving Atomic Charges - The RESP Model. *J. Phys. Chem.* **97**, 10269-10280 (1993).

4

Insulin

Insulin is the most frequently used model system in the study of fibril forming proteins. Aggregation of insulin at the site of injection and in pharmaceutical stocks is both detrimental to health and financially costly. In this chapter the oligomeric population of bovine insulin under aggregating conditions is studied by MS and IM-MS. Multiple distinct conformations of oligomers up to hexameric order are discovered, and their identity confirmed using FT-ICR MS and CID. MD simulations trained by experimental data study the dominant monomeric and dimeric species in atomistic detail.

4.1. Introduction

Insulin is a hormone produced in the β cells of the islets of Langerhans in the pancreas. It is vital for regulating the metabolism, stimulating the conversion of glucose into glycogen and thereby preventing a toxic build up of glucose in the blood stream. After its discovery in 1921 [1] insulin became the subject of intense scientific interest; becoming the first protein to be sequenced in 1955 [2] and its structure determined in one of the earliest successes of crystallography [3]. The biologically active form of insulin is the monomer. An insulin monomer consists of two chains; acidic chain A (21 amino acids) and basic chain B (30 amino acids) which are cross-linked by two disulphide bridges, with an additional intra-chain disulphide bond present in chain A [3] (Figure 4.1).

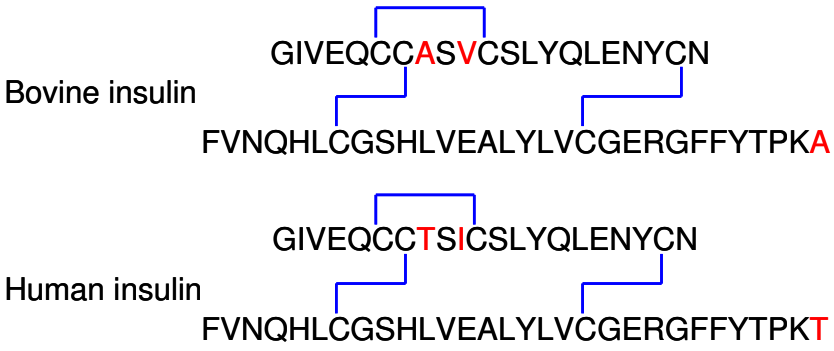


Figure 4.1 | Amino acid sequences of bovine and human insulin. Disulphide bonds and shown in blue and differences in sequence are shown in red.

Insulin is stored as a hexamer, stabilised by the presence of $zinc^{2+}$ ions (Figure 4.2). The regulation of insulin hexamer to monomer conversion is an important homeostatic mechanism.

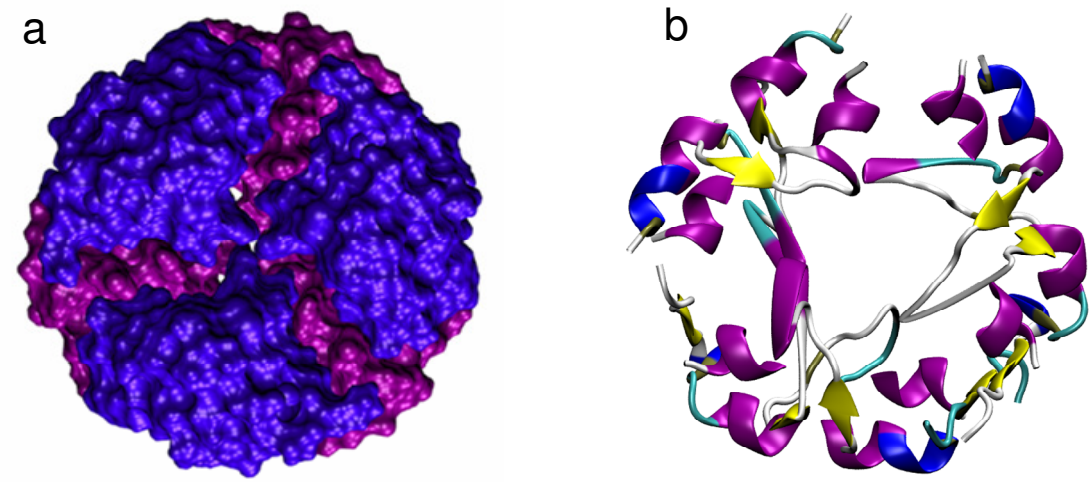


Figure 4.2 | Insulin hexameric structures at pH 6.8 a. Surface representation with monomeric subunits closer to the viewer depicted blue and further away units in purple b. Secondary structure representation: α helix – purple, 3_{10} helix – blue, extended β structure – yellow, turn – cyan, coil – white. Representation rendered using VMD from pdb file 2ZP6 [4].

Insulin hexamers are assembled from three dimers in which the nine N-terminal B chain residues can adopt either a helical conformation (R) or an extended state (T). This creates hexamers of the form T_6, T_3R_3 and R_6 which exist in an allosteric equilibrium [5]. Three histidine residues at position 10 of the B chain coordinate the zinc ions in both conformations. In the R state the tetrahedral coordination of Zn^{2+} requires additionally a ligand, whilst T state Zn^{2+} adopts an octahedral geometry with an additional three water molecules. The conformational change between T and R states is caused by phenolic ligands binding in six hydrophobic pockets, of import as phenolic preservatives are often used in drug preparations. ESI MS can identify the stoichiometric changes which occur between T and R insulin hexamer conformations confirming that these different structures can be preserved into the gas-phase; additionally insulin tetramer, hexamer and decamer oligomers were observed to be present in the mass spectrum in the absence of zinc [6].

Type 1 diabetes results from autoimmune damage to the β cells of the islets of Langerhans, causing a deficit in insulin production [7]. This can be treated by insulin

replacement therapy, requiring careful monitoring of blood glucose levels and daily injections of insulin [7]. Type 2 diabetes often occurs later in life when sufficient insulin cannot be produced to maintain blood glucose levels, or insulin is produced but not used effectively by the body [8]. Injection of monomeric insulin provides the fastest relief for diabetes patients; however monomeric insulin is prone to aggregation, forming fibrils and precipitating out of solution, resulting in a less effective pharmaceutical preparation [9]. Insulin $2Zn^{2+}$ hexamers are more stable [10,11], but after injection are hindered from absorption into the blood stream by their large size [9]. Corresponding patient blood sugar level response is therefore slower as there is a delay whilst hexamers are broken down firstly into dimers, with subsequent dimer dissociation producing monomeric insulin which can reach tissues [9]. Research into a faster acting form of insulin which would be less vulnerable to aggregation led to the development of monomeric insulin analogues [12], which are used extensively in modern diabetes treatment plans [13].

Insulin has been the focus of a number of studies concerning protein self assembly and the resulting fibrillar aggregates [14]. Insulin is a predominantly α-helical protein in its native state, forming a hexamer at neutral pH. Insulin aggregation *in vivo* occurs at the site of insulin injection [15]. Waugh observed insulin fibrils *in vitro* in 1946 and described the sigmoidal kinetics characteristic of amyloidogenesis [16]. The initial lag phase followed by a period of accelerated growth and culminating in saturation previously was ascribed to a nucleation and growth process similar to that observed in crystallisation [17]. However, recent investigations by Knowles *et al.* [18] suggest that primary nucleation is more rapid than previously thought, and that the characteristic sigmoidal growth curve is a result of secondary processes, including fragmentation. Burke and Rougvie in 1972 [19] observed that insulin fibrils displayed the distinctive cross-β X-ray diffraction pattern characteristic of amyloid fibrils [20]. If the amino acid sequence is depleted at the C-terminus of the B chain, insulin becomes more amyloidogenic, whereas joining the N-terminus of the A chain to the C-terminus of the B chain reduces aggregation propensity. These experiments, supported by HDX data showing C-terminal flexibility [21], imply that the C-terminus of the B chain has a vital role in insulin aggregation [11].

Conditions which promote rapid fibrillogenesis *in vitro* are high protein concentrations, low pH (1.5 - 2.0) and temperatures of 60 to 70 °C [11,16,22], all of which favour monomerisation and partial unfolding of the protein. Under aggregating conditions Nettleton *et al.* observed bovine insulin oligomers up to dodecameric order using n-ESI MS, where the population of large oligomers decreased throughout the lag phase [14]. Concurrent EM measurements showed a transition with time from amorphous aggregates to twisted un-branched fibrils containing protofilaments, whilst FTIR revealed a simultaneous increase in the proportion of β structure present [14]. Further FTIR and CD experiments confirmed an initial α-helical structure which converted to β structure as fibrils become visible by EM [23].

During the early stages of self assembly (the apparent "lag" phase) it is predicted that a wide variety of different oligomeric species will be present, all potentially on-pathway to self assembly [18]. In this chapter significantly populated multiple oligomer populations of insulin present in the lag phase prior to fibril assembly are observed using IM-MS, giving new insight into the self assembly of insulin. These oligomers are unequivocally assigned using high resolution FT-ICR MS and CID experiments. Experimental CCS data is used to train MD simulations providing atomistic detail for the dominant dimeric species. The stability of several representative dimers is evaluated *via* binding energy calculations. The association of monomers must be the first step on any self assembly pathway and therefore studies into this interface are vital. This also might enable targeting of this interface by small molecule inhibitors to prevent amyloid formation.

4.2. Methodology

4.2.1. Sample Preparation

Bovine insulin (CAS number: 11070-73-8, Sigma Reference: I5500) was obtained as a lyophilised powder from Sigma Aldrich (UK), with a theoretical mass of 5733.49 Da. The zinc content of the powder was approximately 0.5 %. The bulk of experiments were conducted by creating an aggregating solution of 3 mg/ml (523 μM) bovine insulin by dissolving the protein powder in H_2O adjusted to pH 2 with formic acid. Thioflavin T fluorescence measurements confirm that insulin amyloid fibrils are formed under these solution conditions. These samples were immediately flash frozen in liquid nitrogen and stored in a -28 $^{\circ}$C freezer. Where experiments were performed at different concentrations or in different solution conditions, the sample preparation methodology was similar to that described above. Deconvolution of the mass spectra revealed an experimental bovine insulin monomeric mass of 5733.24 Da. The difference between theoretical and experiment masses of 0.25 Da is most probably due to calibration inaccuracies.

4.2.2. Mass Spectrometry

High resolution mass spectrometry was performed on a QTOF 2 mass spectrometer (Waters, Manchester, UK) and QTOF Ultima mass spectrometer (Waters, Manchester, UK). Mass spectra were recorded for samples in positive mode utilising an n-ESI ionisation source. The n-ESI source was modified to contain a heating block (described in Chapter Two, Section 2.2.1.2) which heated the glass capillary containing the sample prior to ionisation.

4.2.3. Ion Mobility Mass Spectrometry

IM-MS measurements were performed on the ' MoQTOF' [24]. Samples were ionised using an n-ESI source in positive mode with an applied capillary voltage of ~1.8 kV and a cone voltage of 50 V. Transmission of the oligomeric species required an elevated pressure in the source region as reported previously [25]. Ions were accelerated through a potential difference of 39.7 V (unless otherwise stated) and injected into the drift cell. The drift cell was filled with helium gas at a pressure of between 3.0 and 3.5 Torr and a temperature between 298 and 303 K. The temperature and pressure of helium were recorded at the start and end of each mobility scan and the average taken. The voltage applied across the cell was varied from 60 V to 15 V and ATDs recorded at eight drift voltages.

The resolution of the MoQTOF is not sufficient to give insulin ^{13}C spacing. Therefore oligomeric species were assigned using zinc adduct spacing from the mass spectrum in addition to information gleaned from ATDs, CID and FT-ICR MS data.

4.2.4. Collision Induced Dissociation Experiments

Collision induced dissociation (CID) experiments in an argon filled collision cell were performed using a QTOF Ultima (Waters, Manchester, UK). Species were selected by their *m/z* value and accelerated into the collision cell at a variety of voltages. The kinetic energy of the ions entering the collision cell was increased by raising the collision voltage parameter until the signal from the parent ion had been completely lost and only product ions remained.

4.2.5. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

High resolution MS was performed on a 9.4 Tesla Apex Qe Fourier Transform Ion Cyclotron Resonance mass spectrometer (Bruker Daltonik GmbH). This provided ^{13}C isotopic distributions which confirmed the identification of peaks as isobaric aggromers or as conformers of the same species. The n-ESI source from a QTOF instrument was clamped in front of the capillary inlet and an external voltage applied to produce spray (for further details please refer to Chapter 2). The source accumulation time was 0.5 seconds; an excitation frequency sweep from 24055.269 Hz to 161282.396 Hz was applied prior to detection from *m/z* 895 to 6000. The voltages applied to the accumulation hexapole were 8.5 V on the entrance and 12 V on the exit. Data acquisition size was 1,048,576 bytes.

Calibration of 6 to 10 ppm accuracy was achieved using Agilent Technologies ESI tune mix (catalogue number G2421A) and data processed using DataAnalysis 4.0 (Bruker Daltonik GmbH). The simulated isotopic distributions were created from theoretical empirical formulas using the Simulate Isotopic Pattern function of DataAnalysis 4.0 (Bruker Daltonik GmbH). Isotopic patterns confirmed that all disulphide bridges in the insulin monomers, and in oligomer constituent monomers, were intact.

4.2.6. Simulation Methodology

All simulations were performed by Massimiliano Porrini.

4.2.6.1. Monomeric Species $[M+3H]^{3+}$ and $[M+4H]^{4+}$

Insulin would be too computationally expensive to simulate with an explicit solvent model, therefore solvent was represented with a continuum solvation method, termed

"OBC" [26,27]. The temperature control at 70 °C was implemented using the Langevin algorithm, with a collision frequency equal to 1.0 ps^{-1}. All the bonds involving hydrogen atoms were constrained at their equilibrium value using the SHAKE algorithm [28], allowing the utilisation of a 2 fs time step.

4.2.6.2. Correlation Between CCS and R_g

The radius of gyration (R_g) and theoretical CCS were calculated for 500 structures of $[M+3H]^{3+}$ to investigate whether there is a correlation between these two variables. To generate 500 structures of $[M+3H]^{3+}$ an implemented loop scheme was used and is detailed below.

DO i = 1, 500
heating from 0 K to 800 K in 6.4 ps
dynamics at 800 K for 60 ps
gradual exponential stepwise cooling from 800 K to 0 K, 2 ps per step
minimisation
END DO

4.2.6.3. MM-PBSA Calculations

To perform Molecular Mechanic - Poisson-Boltzman Surface Area (MM-PBSA) calculations dimers were immersed in a TIP3P [29] water molecules box of approximately 80 x 70 x 70 $Å^3$ (varying with the dimer under investigation) and containing between ~ 9500 and ~ 11500 water molecules. The system's +6, +7 and +8 charge was neutralized with 6, 7 and 8 chloride ions respectively. After equilibrating temperature (70 °C) and density at 1 atm of pressure, a production simulation of up to 25 ns in NPT ensemble was run, with temperature collision frequency of 2.0 ps^{-1} and a pressure relaxation time of 2.0 ps. All of the bonds involving hydrogen atoms were constrained at their equilibrium values, so that a time step of 2.0 fs could be used. To derive the electrostatic interactions a particle mesh Ewald (PME) method [30] was implemented, using a radial cut off of 8.0 Å.

4.3. Results and Discussions

4.3.1. Mass Spectrometry

Mass spectrometry oligomer populations reflect those present in solution [31]. Conditions which promote insulin fibrillogenesis are high protein concentrations, low pH (1.5 - 2.0) and temperatures of 60 to 70 °C [11,16,22]. However, the precise conditions employed affect the rate of fibrillogenesis and the morphology of fibrils produced. For example, different acids produce different rates of insulin fibrillisation [32] and different mature fibril morphologies [33]. MS and IM-MS experiments were undertaken in a variety of solution conditions, to assess the impact of protein concentration, solution pH and acid type on oligomer populations.

4.3.1.1. Different Acids

Experiments were undertaken with both formic (HCOOH) and hydrochloric (HCl) acid solution conditions, at a variety of insulin concentrations. Figure 4.3 is a comparison of the mass spectra produced when similar concentrations of insulin were prepared in aqueous formic and hydrochloric acids at pHs of 2.0 and 1.6 respectively.

Under both conditions the dominant spectral peak in Figure 4.3 can be assigned as $[M+4H]^{4+}$. Despite the greater acidity of the hydrochloric acid solution, for formic acid the monomeric charge envelop includes both $[M+5H]^{5+}$ and $[M+6H]^{6+}$ species at significant abundance. This may be because although more extended monomeric states are of greater prevalence under the more acidic HCl solution conditions, heavy adduction of insulin monomers by Cl^- ions reduces the net positive charge monomers carry and are detected with. Another possibility is that acid-induced refolding of insulin is occurring in the HCl solution [34].

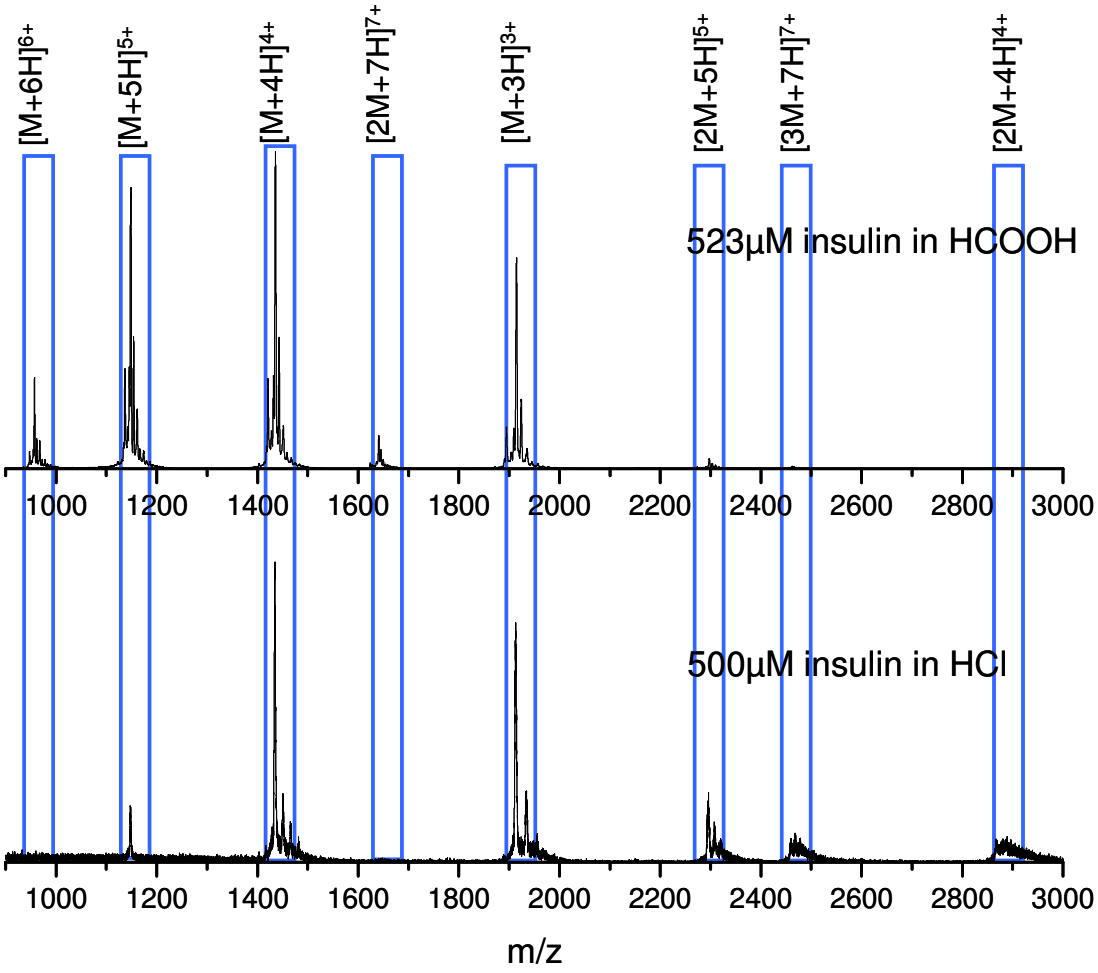


Figure 4.3 | Bovine insulin in aqueous formic (pH 2) and hydrochloric acid (pH 1.6) at similar concentrations.

Although the intensity of the higher order oligomers appears greater in hydrochloric acid, this is misleading as the quality of mass spectra are inferior, as visible in the raised baseline visible at the lower *m/z* values. Additionally, due to the repeated adduction of oligomers by chloride ions, the spectrum resolution is substantially degraded. Chloride adducts could be removed by increasing the cone voltage applied to the source region of the mass spectrometer, leading to a more disruptive desolvation process. However, this caused fragmentation of the higher order oligomers, thereby negating it as a strategy. Conversely, aqueous formic acid solution conditions generated a wider array of oligomeric species, both in terms of

charge states and multimeric order, with substantially better resolution and fewer adducts.

4.3.1.2. Different Protein Concentrations

Insulin was prepared in a range of different concentrations in both hydrochloric and formic acid aqueous solutions.

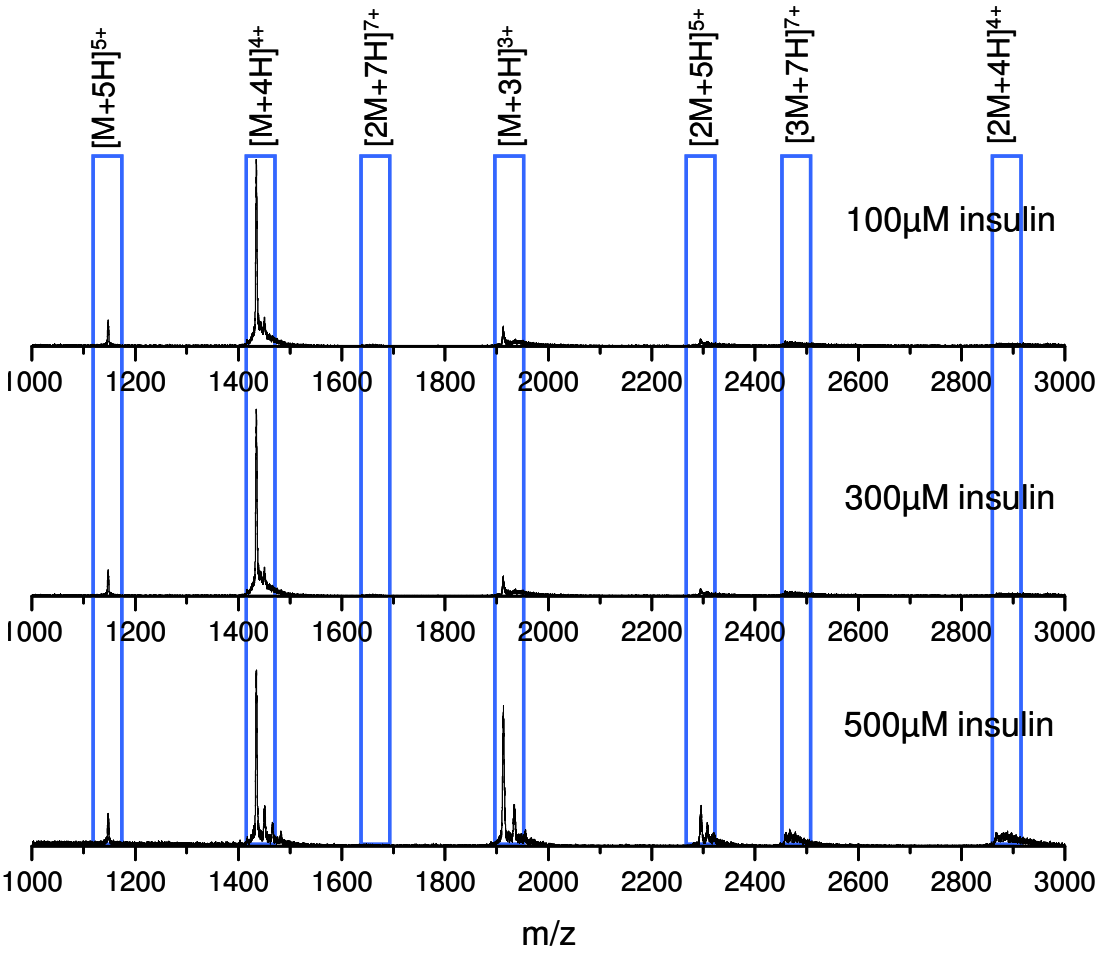


Figure 4.4 | Bovine insulin at three concentrations, in aqueous hydrochloric acid at pH 1.6.

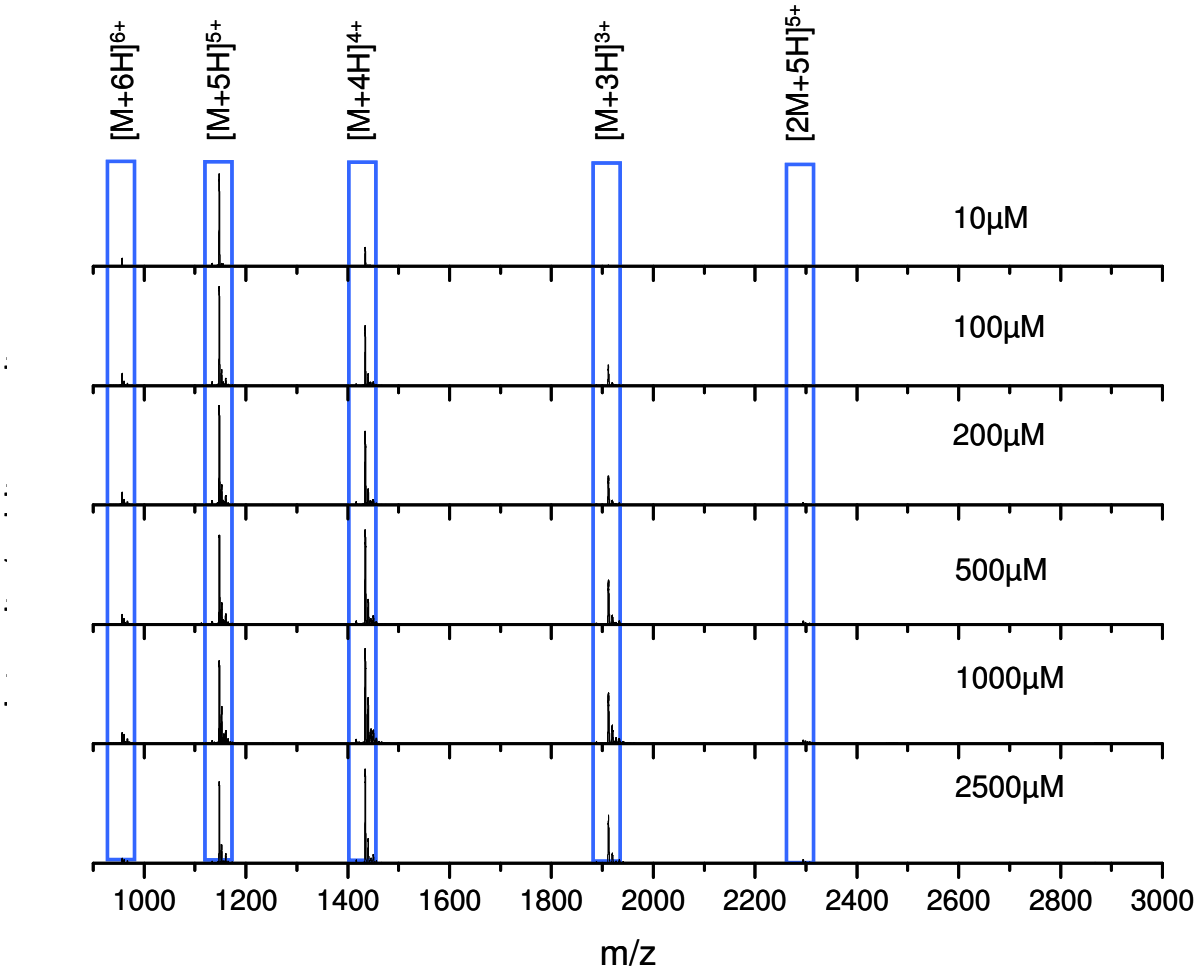


Figure 4.5 | Spectra of insulin in aqueous formic acid at various concentrations.

The dominant species in the majority of spectra is $[M+4H]^{4+}$. An exception to this is for the lower concentrations of insulin in aqueous formic acid where $[M+5H]^{5+}$ is of the highest intensity. At higher concentrations, independent of acid type, there is an increase the number and population of larger oligomers observed. This could provide an explanation for the dominance of low charge state monomers at higher concentrations, as these could result from the dissociation of larger oligomers. An enlargement of the higher *m/z* region for formic acid solution conditions allows greater discernment (Figure 4.6).

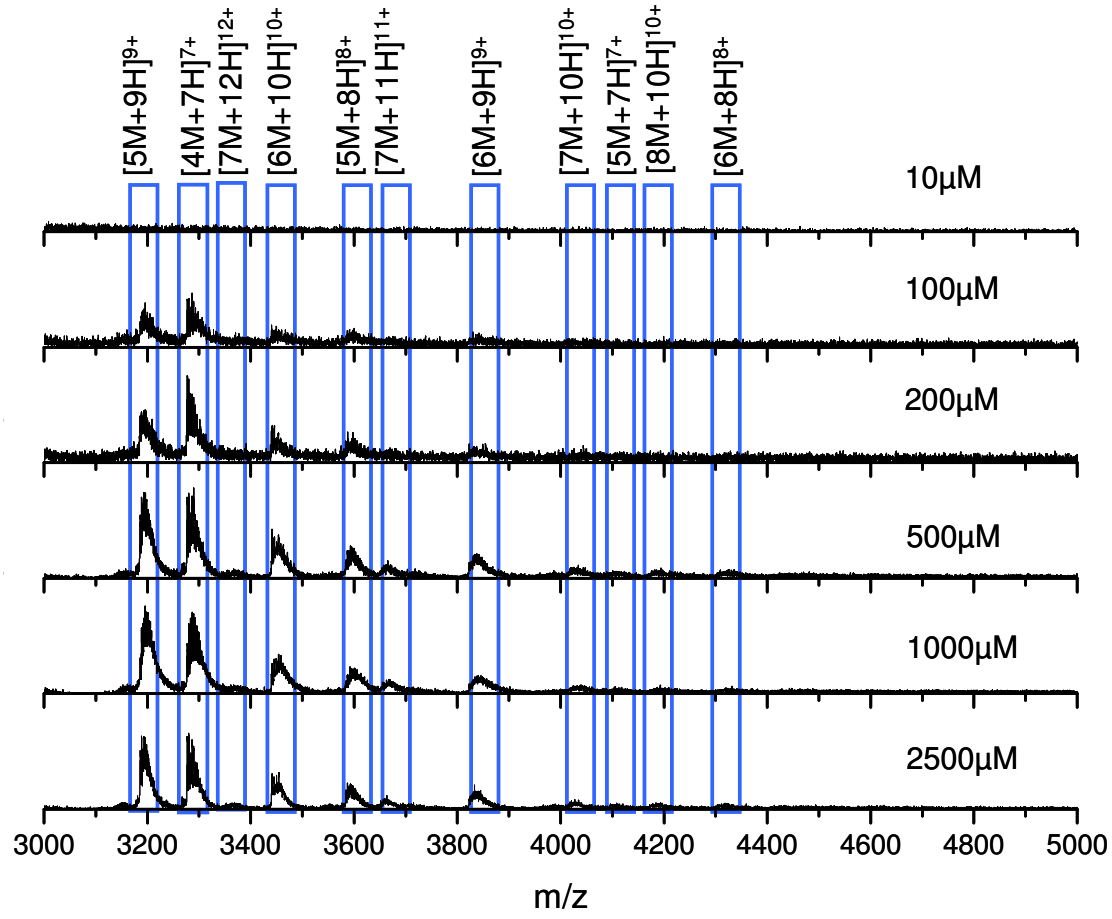


Figure 4.6 | Enlargement of the high m/z region of the spectra of insulin in aqueous formic acid at various concentrations.

At 500 µM and above a wide range of larger oligomers are present at significant intensities. Higher order oligomer populations remain quantitatively similar with increasing concentrations after a threshold of 500 µM has been reached. If the formation of oligomers was a n-ESI effect no such threshold would exist. Therefore a solution of aqueous formic acid containing insulin at a concentration of 523 µM was chosen for the majority of further experiments. This decision was made based on the higher quality spectra formic acid afforded, coupled with the abundant range of oligomers displayed at this concentration. It should be noted that IM-MS measurements of oligomers formed in a solution of aqueous hydrochloric acid showed very similar CCSs to those formed in formic acid (data not shown). For

clarity only CCSs of insulin oligomers formed in formic acid solutions are discussed later in this chapter.

4.3.1.3. Different pHs

To obtain solutions of different pHs a stock solution of bovine insulin in aqueous formic acid was diluted with water, ammonium acetate and acetic acid in varying quantities, to create three samples of 40 µM concentration at pHs of 2, 4.7 and 5.5.

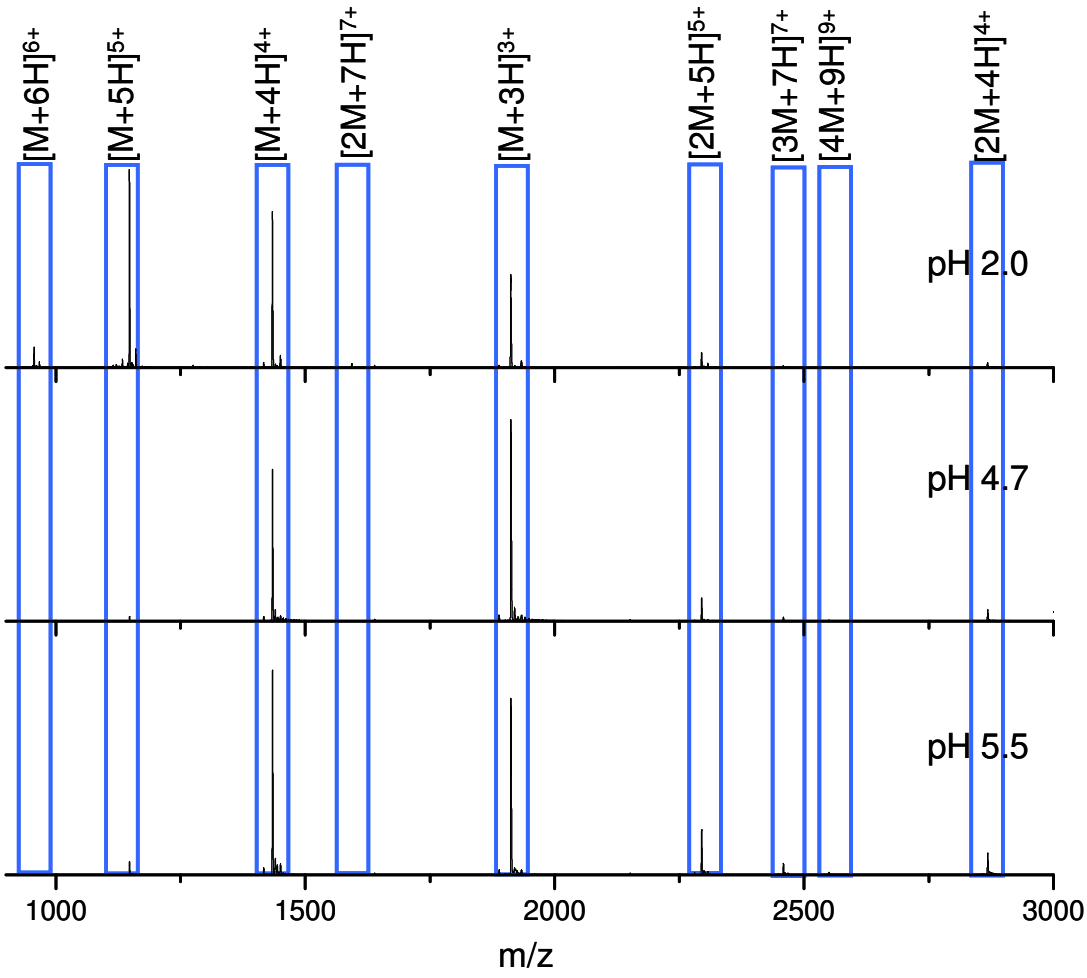


Figure 4.7 | Mass spectra of 40 µM insulin at varying pHs.

Surprisingly, results obtained were at odds with previous publications; whilst Nettleton *et al.* report dominant $[6M+11H]^{11+}$ and $[6M+10H]^{10+}$ species at pH 4 [14], in our experiments monomeric species dominate the mass spectra at all pHs employed, although hexameric species are present. We speculate that this might be due to the lower insulin concentrations employed in this experiment as Nettleton *et al.* utilise 2 mM concentrations. In line with expectations, a pH 2 solution produces a greater average charge per monomer than more basic solutions. However, more unusually, pH 5.5 has a dominant $[M+4H]^{4+}$ species whilst at pH 4.7 the $[M+3H]^{3+}$ species is in greater abundance.

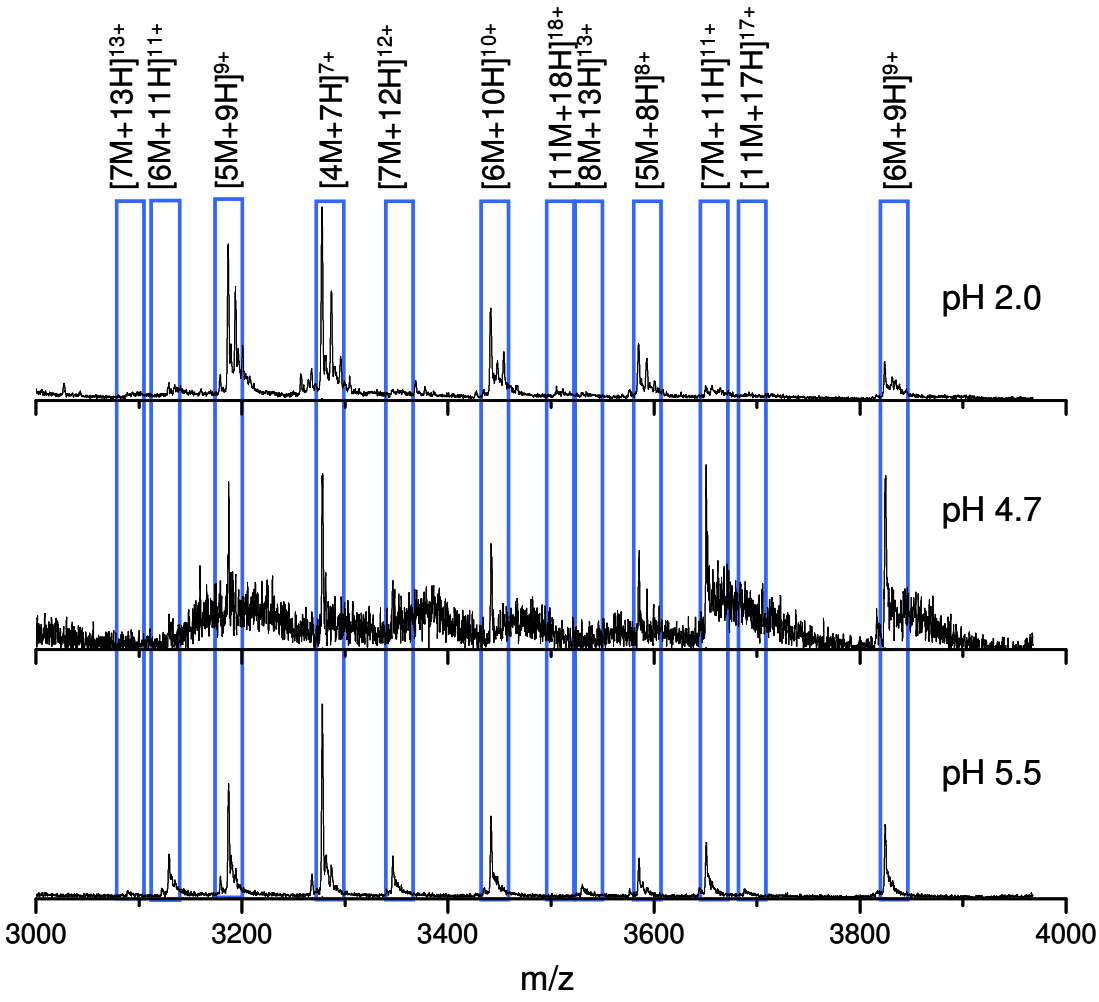


Figure 4.8 | Enlargement of the high m/z region of the mass spectra of 40 µM insulin at varying pHs. The poor quality of the spectrum at pH 4.7 is due to increased salt adduct presence.

The above figure displays an enlargement of the higher *m/z* region. At pH 5.5 all hexameric and heptameric species are more abundant than in more acidic solutions. pH 5.5 possibly favours the formation of hexamers in solution as it mimics the physiological conditions found in the secretory vesicles where insulin is formed [35]. The preponderance of heptamers is more perplexing, as they are of no known relevance *in vivo*. It may be a consequence of the increased hexamer quantity, as monomer addition to these would create greater numbers of heptamers than might otherwise be expected. This would suggest a very dynamic system, with unstructured hexamers available for monomer addition. At pH 2, where aggregation of insulin is known to occur, a greater range of species and more high mass oligomers are observed.

4.3.1.4. Selected Conditions

Bovine insulin was dissolved in water, adjusted to pH 2 with formic acid, to a concentration of 523 μM. At room temperature these solution conditions promote aggregation; however fibrillogenesis takes place over a period of months rather than the hours taken at elevated temperatures [36]. In support of this, the mass spectra and conformer distributions reported below did not vary significantly throughout any 6.5 hour experiment. We have confirmed that fibrils are indeed formed under these conditions by Thioflavin T fluorescence measurements and EM observations (data not shown). The resulting mass spectrum (Figure 4.9) shows a wide variety of species. Oligomers of the general form $[nM+zH]^{z+}$ are observed with $1 \leq n \leq 15$. All of the species observed by n-ESI MS carry more charges than the multimeric order of the oligomer. The largest intensity species are attributed to two monomers of adjacent charge states: $[M+5H]^{5+}$ and $[M+4H]^{4+}$ at *m/z* values 1147 and 1434. However, there is also a significant contribution from peaks at *m/z* values 1639 and 2294, where the principal contribution is from dimeric species (see below). Zinc adducts are present on all peaks, due to the presence of zinc in the insulin preparation. However, the dominant peak of both monomers and oligomers is unadducted by zinc.

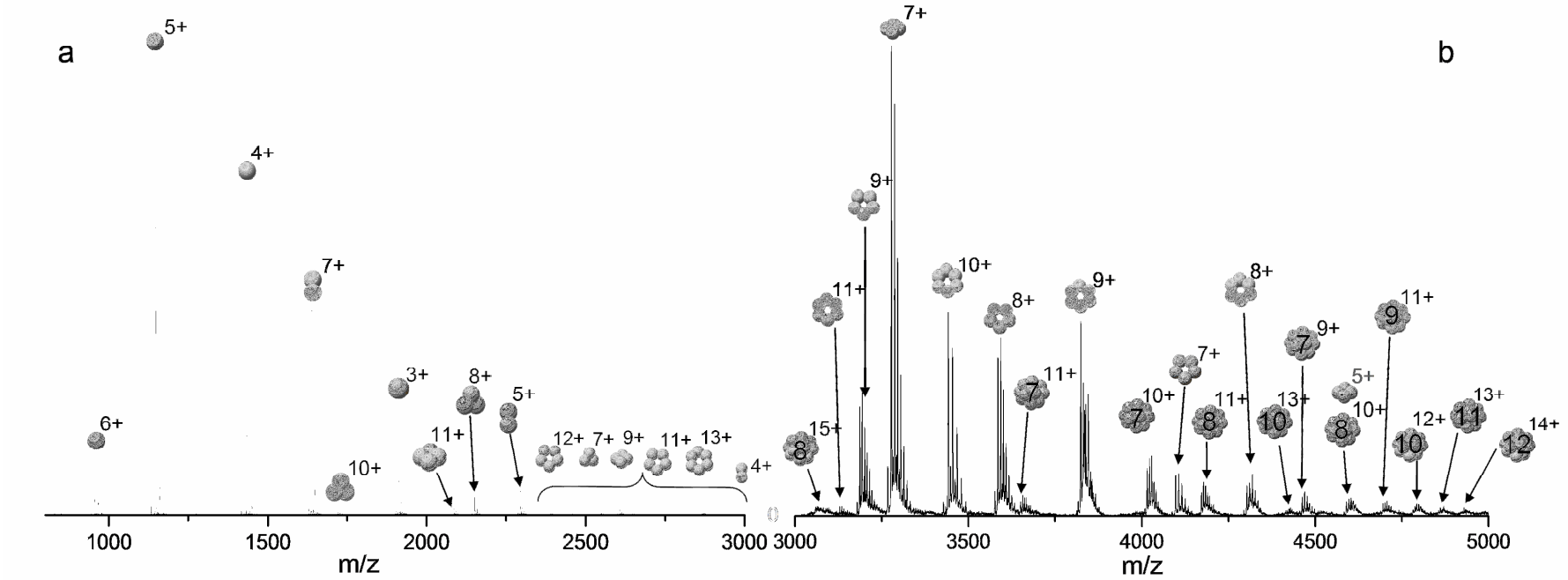


Figure 4.9 | Mass spectra of bovine insulin a. is a typical n-ESI mass spectrum obtained from a solution of bovine insulin. b. shows an enlargement of the higher m/z region.

4.3.1.5. Aggregate Destabilisation

Experiments which destabilise aggregate populations by subjecting them to greater forces, either during desolvation or inside the mass spectrometer, can give valuable insight into oligomer stability and dynamics within the population.

4.3.1.6. Cone Voltage Experiments

As mentioned previously, when the cone voltage is increased, aggregate species are observed to break up.

The dimeric species $[2M+5H]^{5+}$ and $[2M+4H]^{4+}$ increase in abundance as the cone voltage is raised (Figure 4.10). This suggests that the populations of these lowly charged dimers are created, in part, by the fragmentation of larger oligomers which are vulnerable to harsh desolvating conditions. Conversely, the peak corresponding to $[2M+7H]^{7+}$ decreases with the increasing cone voltage, although it is more persistent at high cone voltages than the populations of $[4M+11H]^{11+}$ and $[3M+8H]^{8+}$ which also decline. The most significant variations in the monomer population with increasing cone voltage occur for $[M+3H]^{3+}$, which increases in intensity.

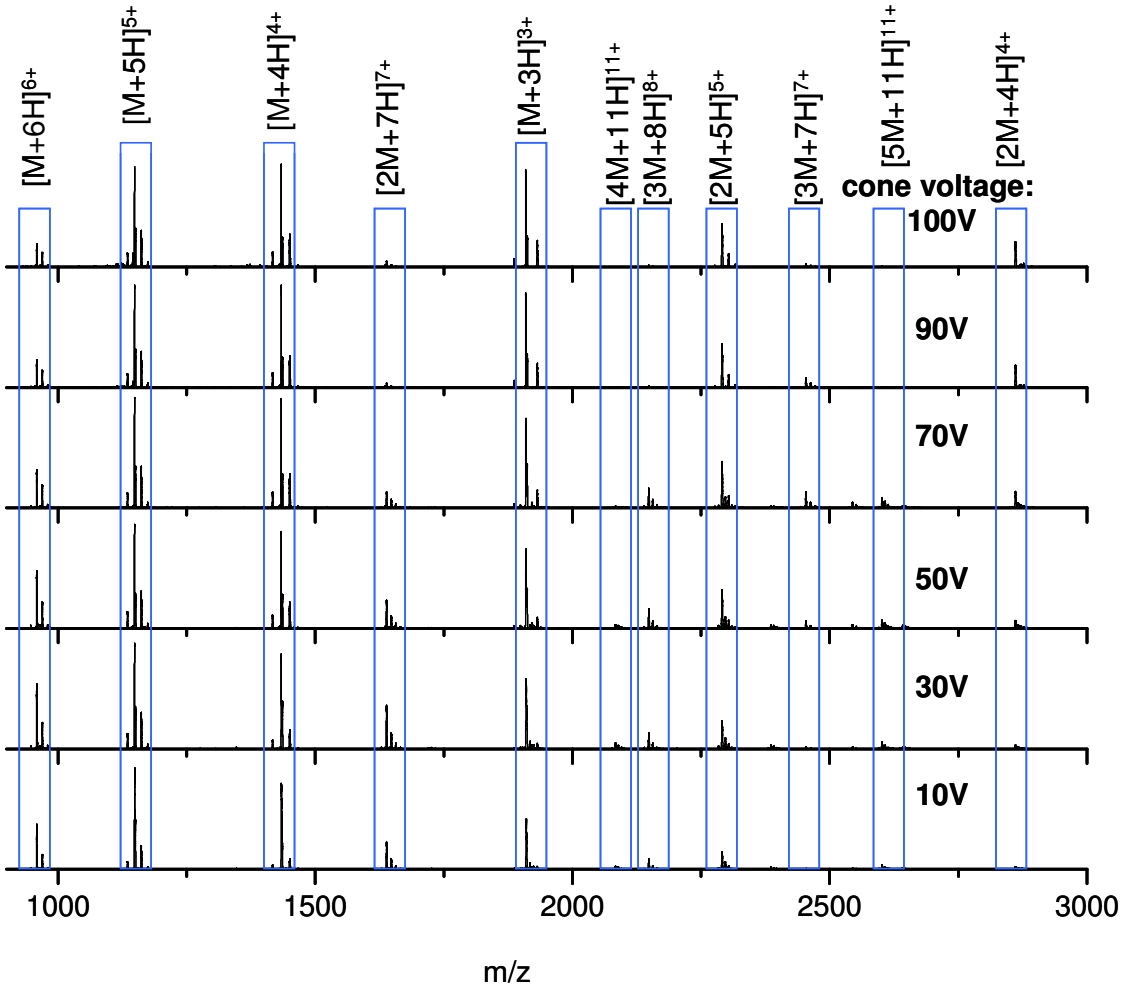


Figure 4.10 | Spectra taken at increasing cone voltages showing oligomer population changes.

CID experiments, described in detail later in the chapter, show that the $[2M+7H]^{7+}$ dissociates into $[M+3H]^{3+}$ and $[M+4H]^{4+}$. Therefore the increase in $[M+3H]^{3+}$ intensity observed in Figure 4.9 could be due to the fragmentation of the $[2M+7H]^{7+}$ ion and resulting repopulation of the $[M+3H]^{3+}$ and $[M+4H]^{4+}$ species.

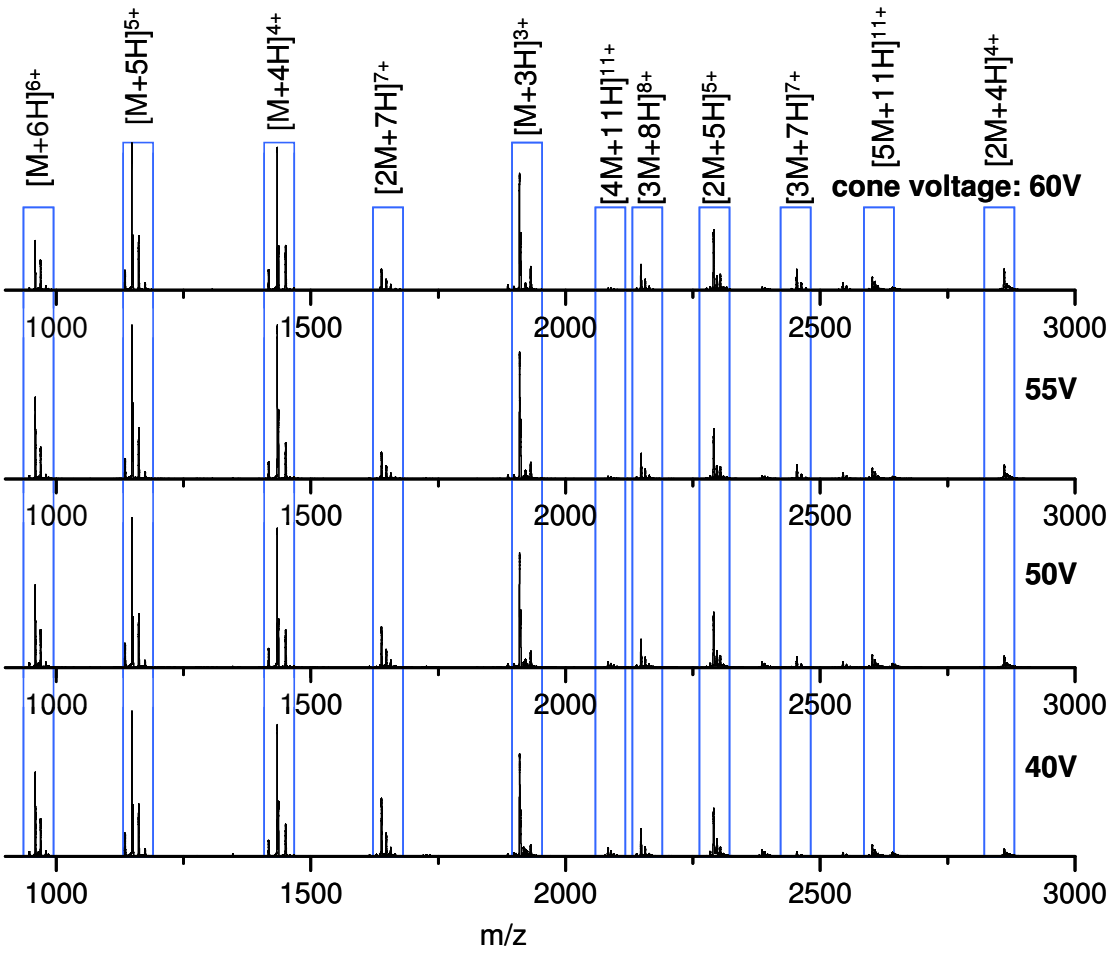


Figure 4.11 | Spectra taken at increasing cone voltages, enlarged to show in detail the region between 40 V and 60 V.

As can be observed from Figure 4.11, no appreciable dissociation of multimers takes place at cone voltages below 60 V. As a cone voltage of 50 V is employed in the experiments which follow, we are operating below the limit at which voltages dissociate oligomers. High voltages cause more energetic collisions between oligomers and the solvent or desolvation gases, however, the voltage must be sufficient to allow oligomer desolvation.

4.3.1.7. Collision Voltage Experiments

In an analogous experiment, raising the energy with which ions enter the collision cell can cause the break up of large oligomeric species into smaller species and the fragmentation of monomers into small peptides. In the figures beneath, there is a clear decrease in the populations of trimers, tetramers and pentamers with increasing collision voltages. Higher collision energies cause more highly charged species to experience a greater acceleration and impact with gas atoms. However, as these ions are travelling faster they will undergo fewer collisions as they travel through the collision cell. These conflicting effects suggest that the persistence of $[M+4H]^{4+}$ and $[M+3H]^{3+}$ visible in Figure 4.12 is most probably due to their repopulation by the break up of $[2M+7H]^{7+}$, rather than that they possess a greater stability than $[M+6H]^{6+}$ or $[M+5H]^{5+}$, which visibly decline with increasing collision voltage.

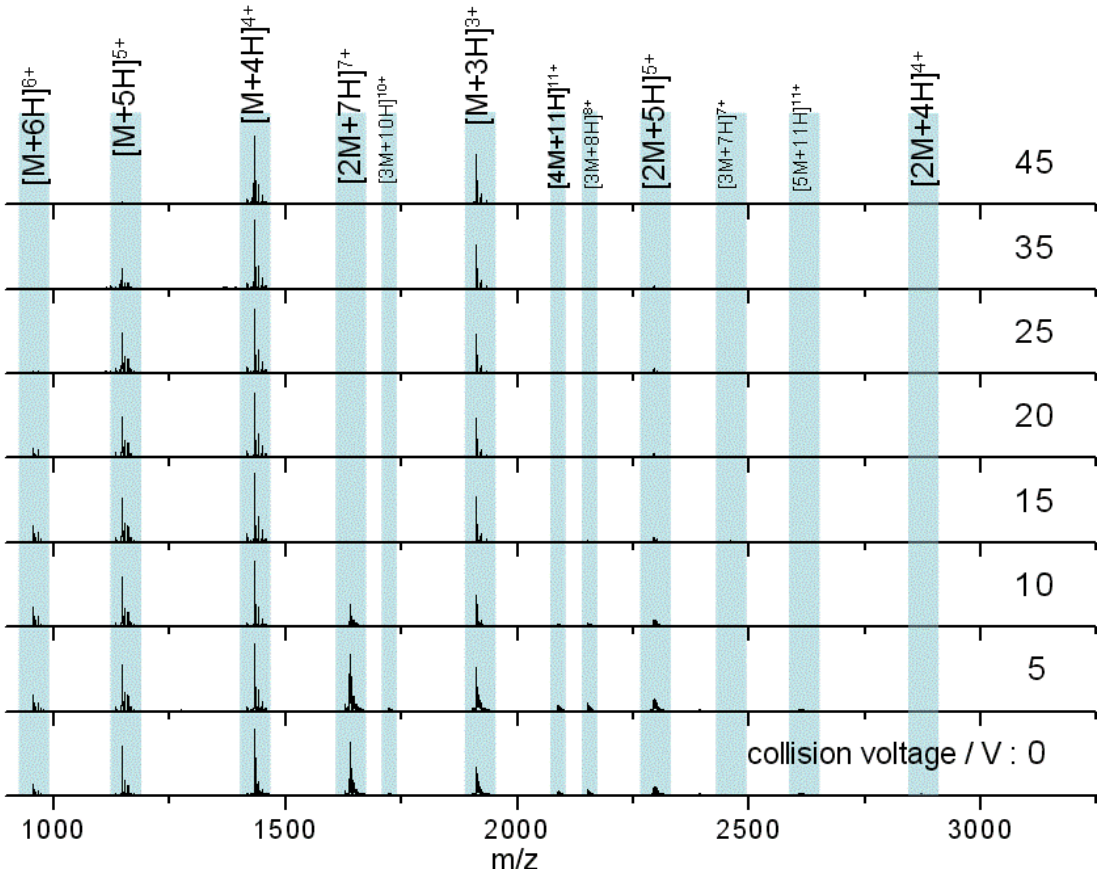


Figure 4.12 | Spectra at increasing collision voltages showing oligomer population changes.

Focusing again on the dimeric species present; $[2M+7H]^{7+}$ is only very sparsely populated when the collision voltage is greater than 10 V (Figure 4.13). Conversely $[2M+5H]^{5+}$ persists with a greater relative abundance at higher collision energies, even at a voltage of 45 V. This is not analogous to the explanation for changing monomer intensities above, as although $[2M+7H]^{7+}$ will experience greater acceleration and more forceful collisions than $[2M+5H]^{5+}$, the charge density of the dimers is considerably less than that of the monomers. There is also an appreciable quantitative increase in the population of $[2M+5H]^{5+}$, a phenomenon unexplained by a less violent collision process.

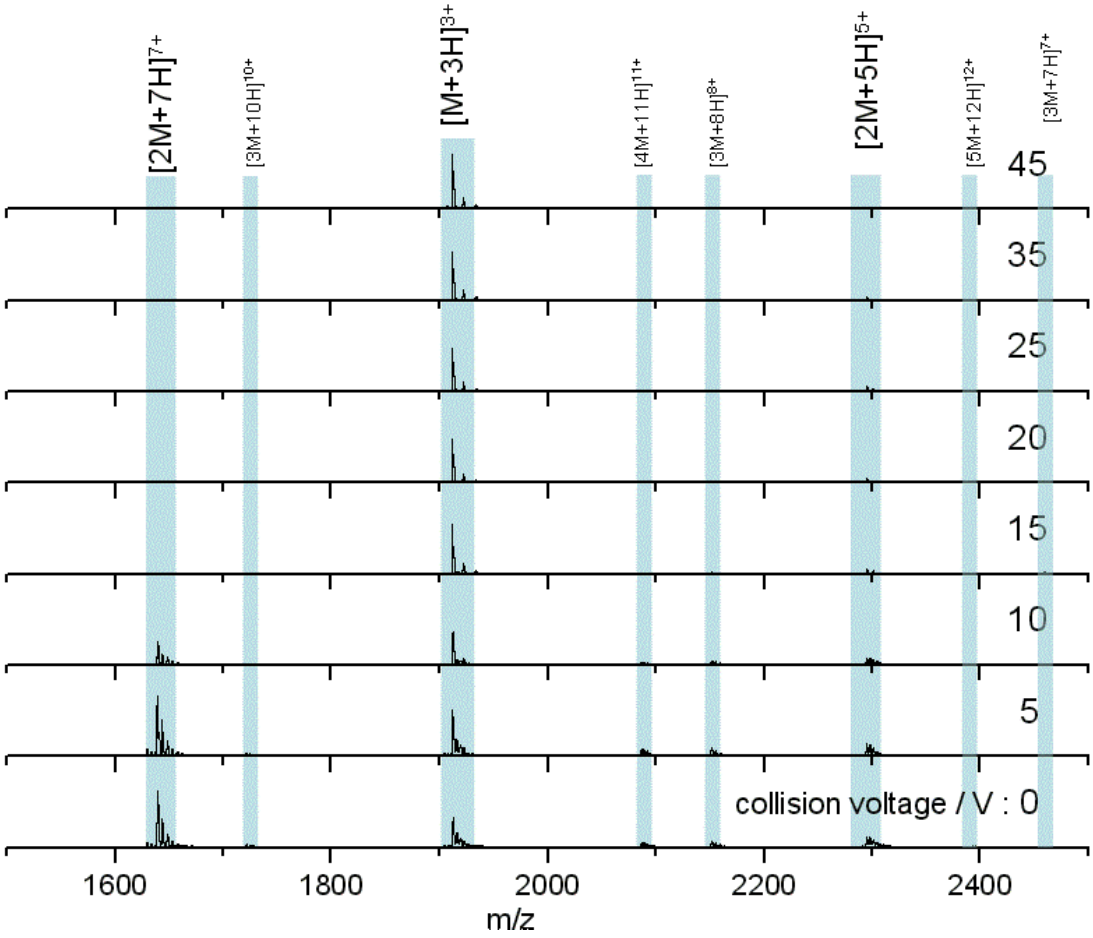


Figure 4.13 | Spectra at increasing collision voltages enlarged to show in detail the m/z region between 1500 and 2500.

Further experiments were performed (data not shown) in which the pressure in the source of the mass spectrometer was lowered. These conditions are observed to break up the higher order oligomers, and identical trends with the decrease of $[2M+7H]^{7+}$ and increase in the relative population of $[2M+5H]^{5+}$ are perceived.

As demonstrated above, under electrospray conditions which best preserve oligomers, a predominant species is $[2M+7H]^{7+}$. When however, experimental conditions favour the break-up of aggregates, we observe an increase in the intensities of species assigned to $[2M+5H]^{5+}$ and $[2M+4H]^{4+}$. It can be inferred that $[2M+7H]^{7+}$ is significantly populated in solution, whereas the lower charge state dimers arise from fragmentation of larger species. Given that the dimer is a stable product ion and also a dominant species in the mass spectrum (Figure 4.9 (a)) we speculate that a dimer may be the core stacking unit in larger aggregates and therefore a key component of the fibrillogenesis process. A proposed schematic of this process is displayed in Figure 4.14.

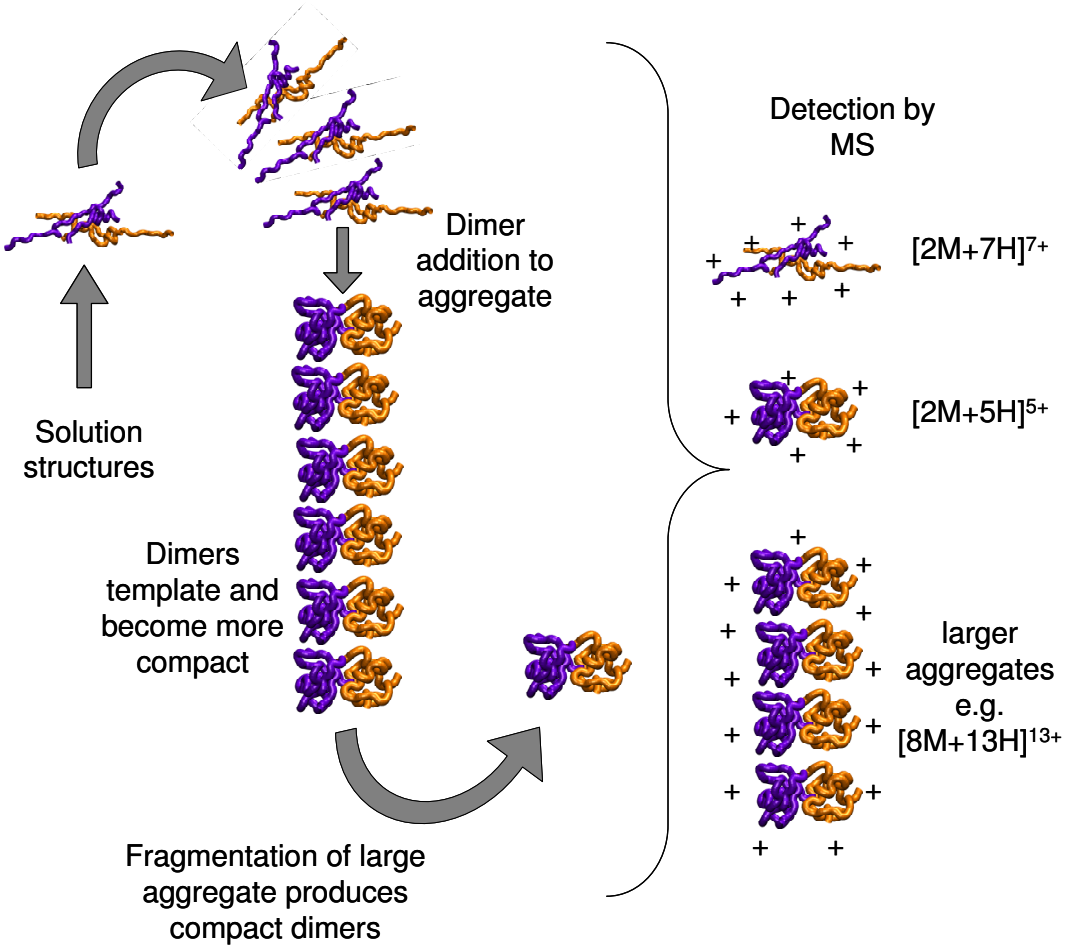


Figure 4.14 | Schematic of a proposed insulin dimer assembly mechanism. $[2M+5H]^{5+}$ and $[2M+7H]^{7+}$ structures are generated by molecular modelling (described later in the chapter) and rendered with VMD software.

4.3.1.8. Time Course Experiments

In order to observe changes in oligomeric populations during insulin aggregation a time course was conducted. An eppendorf of bovine insulin, at a concentration of 523 μM in an aqueous solution acidified to pH 2 with formic acid, was heated to 67 °C in a Jencons digital dry bath. The source heating block, described in detail in Chapter Two, was used to ensure samples were kept at 67 °C, prior to ionisation,

whilst the experiment was conducted. A six hour time course was performed; sample spectra from the time course are shown in Figure 4.15.

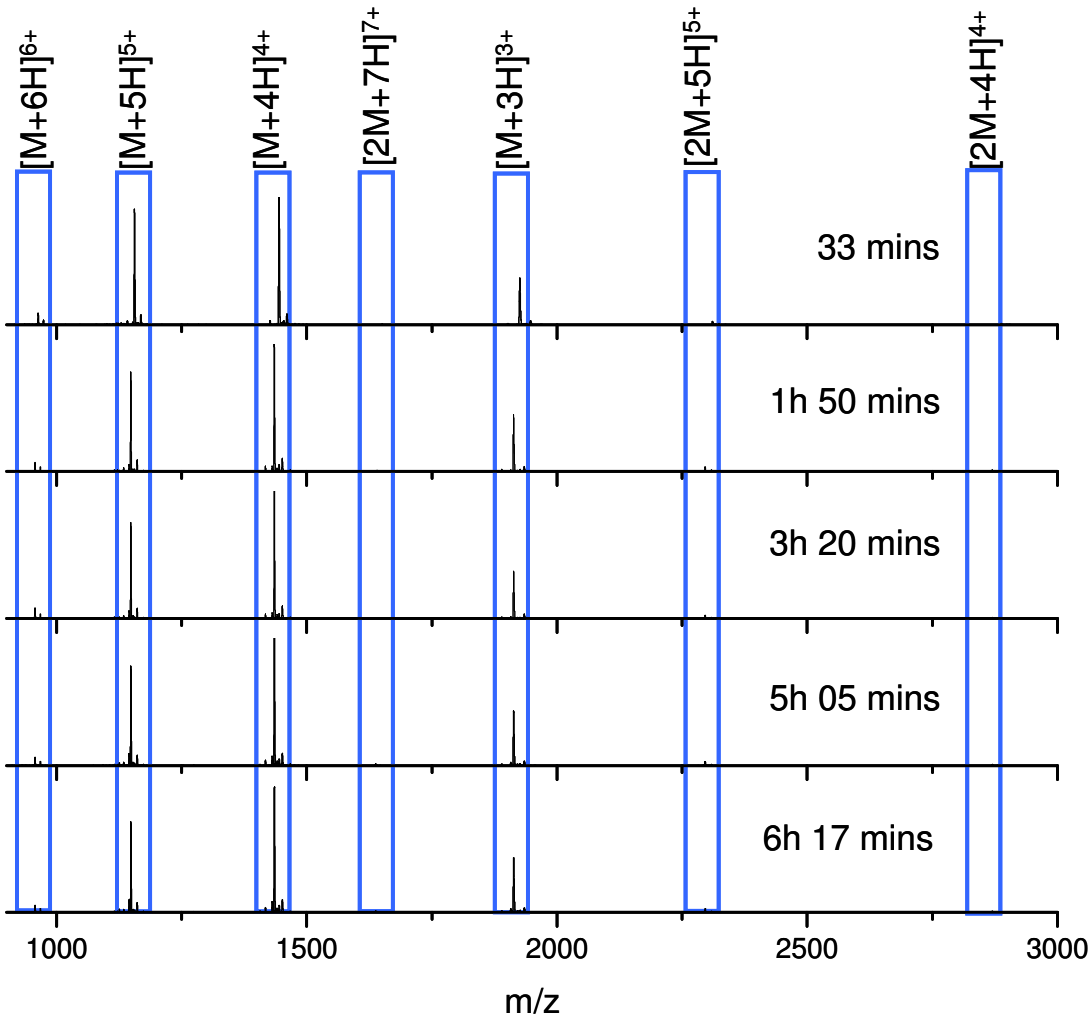


Figure 4.15 | Time course spectra of 523 μM insulin in an aqueous solution acidified to pH 2 with formic acid, heated to 67 °C.

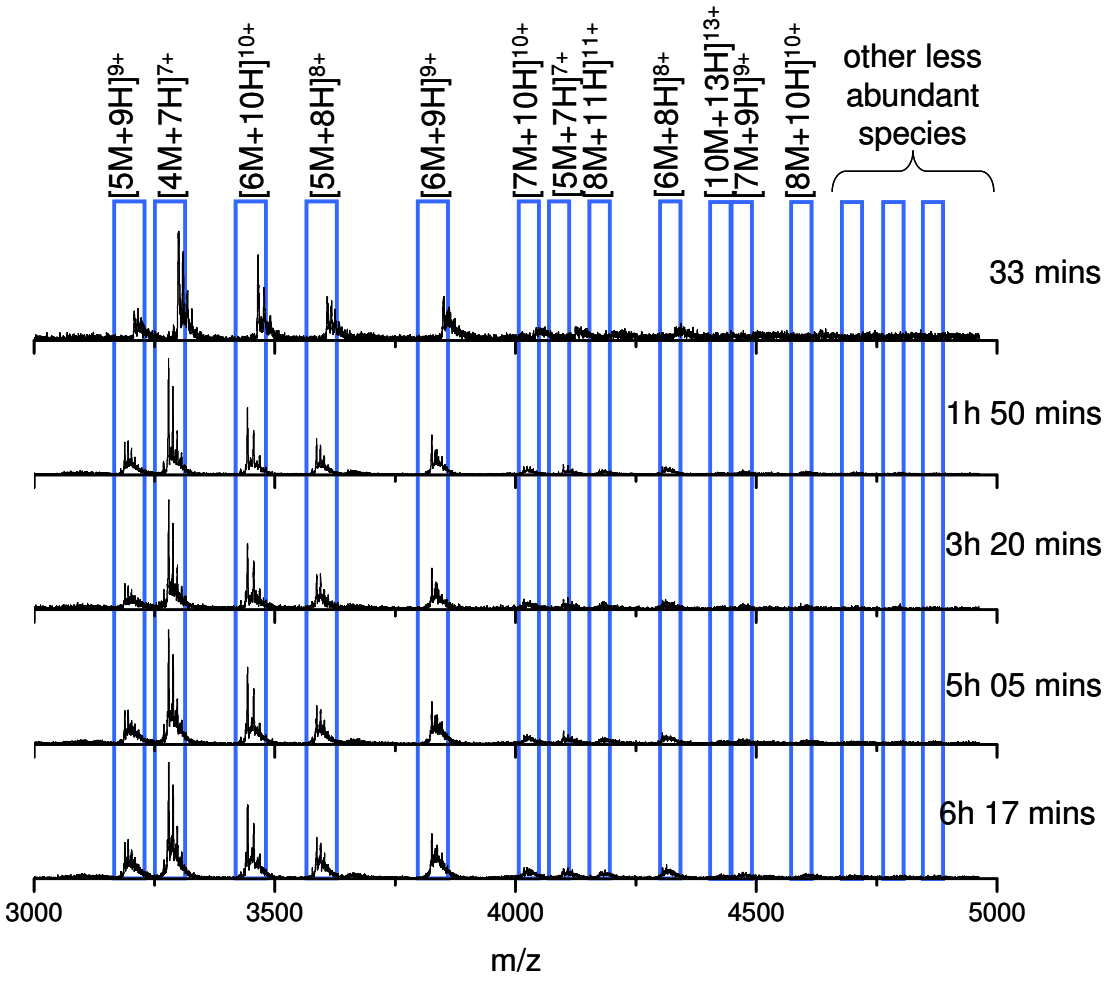


Figure 4.16 | Enlarged higher m/z region of time course spectra of 523 μM insulin in an aqueous solution acidified to pH 2 with formic acid, heated to 67 °C.

No significant changes in the relative oligomer populations occurred over the experimental time course (Figure 4.15 and Figure 4.16). The experimental conditions utilised are conducive to aggregation, which was indirectly observed in the clogging of capillary tips as shown in Figure 4.17. IM-MS was also performed for a two hour time course, under identical solution conditions; no change in oligomer CCSs occurred.

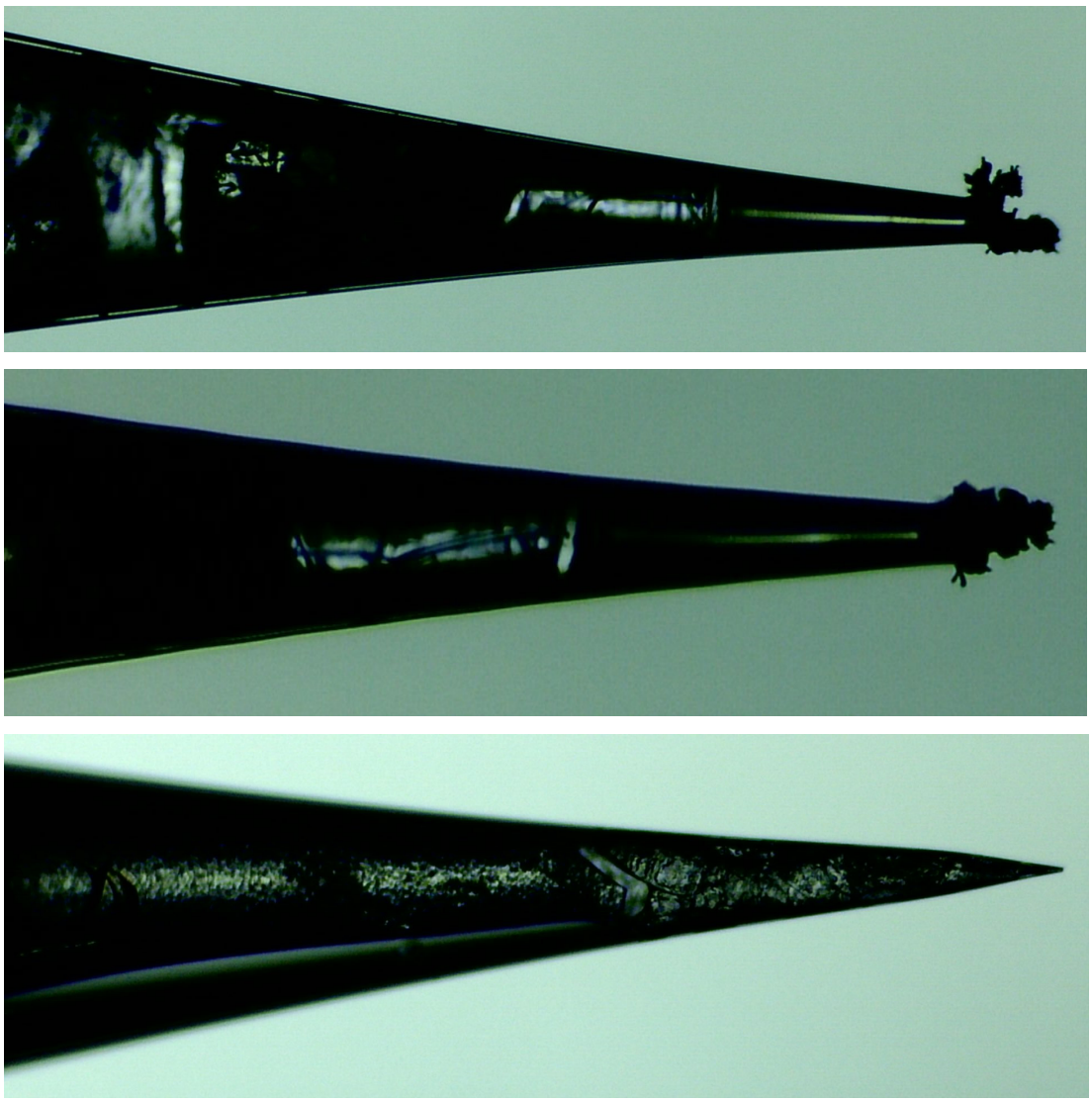


Figure 4.17 | Photographs of capillary tips blocked by aggregating insulin solutions during time course experiments.

4.3.2. Ion Mobility Mass Spectrometry

4.3.2.1. Multiple Conformation Observations

IM-MS has the valuable ability to differentiate between different conformational forms of ions, and also ions which have coincident *m/z* values, by their differing mobilities when travelling through a drift cell. Deconvolution of the ATDs of insulin oligomers $[2M+7H]^{7+}$, $[2M+9H]^{9+}$, $[3M+8H]^{8+}$, $[3M+10H]^{10+}$, $[4M+11H]^{11+}$, $[5M+11H]^{11+}$, $[5M+12H]^{12+}$ and $[6M+13H]^{13+}$ revealed multiple peaks present in the ATDs for all drift voltages. These could correspond to multiple conformations of a single species or oligomers of a coincident *m/z* (aggromers). Bernstein *et al.* recently observed multiple peaks in the ATD of a $[2M-5H]^{5-}$ Aβ amyloidogenic peptide species; they postulated that these corresponded to higher order hexamer aggromers, which stacked to form a dodecameric paranucleus [37]. A stacking mechanism would validate the assignation of the additional ATD peaks we observe experimentally to much larger oligomers, without the need for the presence of intermediate order oligomers which might otherwise be expected.

To validate assignment of ATD peaks as either conformers of a species, or *m/z* coincident species, high resolution FT-ICR MS and CID are utilised. Figure 4.18, Figure 4.19, Figure 4.20 and Figure 4.21 illustrate the use of IM-MS, CID and FT-ICR MS to determine unequivocally ATD peak species identity.

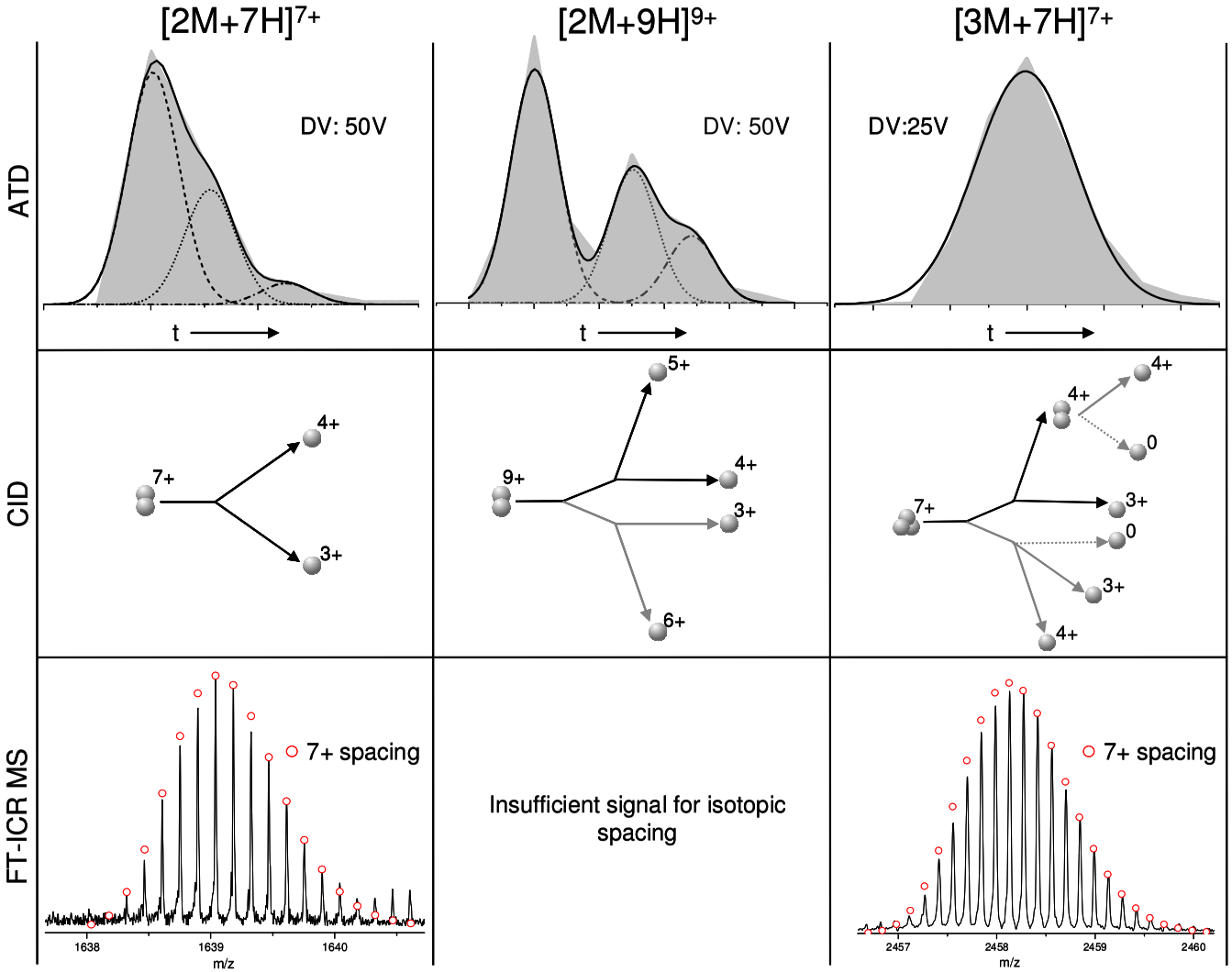


Figure 4.18 | ATD, FT-ICR MS and CID data for $[2M+7H]^{7+}$, $[2M+9H]^{9+}$ and $[3M+7H]^{7+}$.

For ATD data the experimental data are the shaded grey area. Dotted and dashed lines represent the Gaussian fits for each species. The solid black line is the cumulative fit of the Gaussian curves.

For CID data black lines denote the dominant fragmentation pathway, dotted lines represent species which cannot be observed by MS.

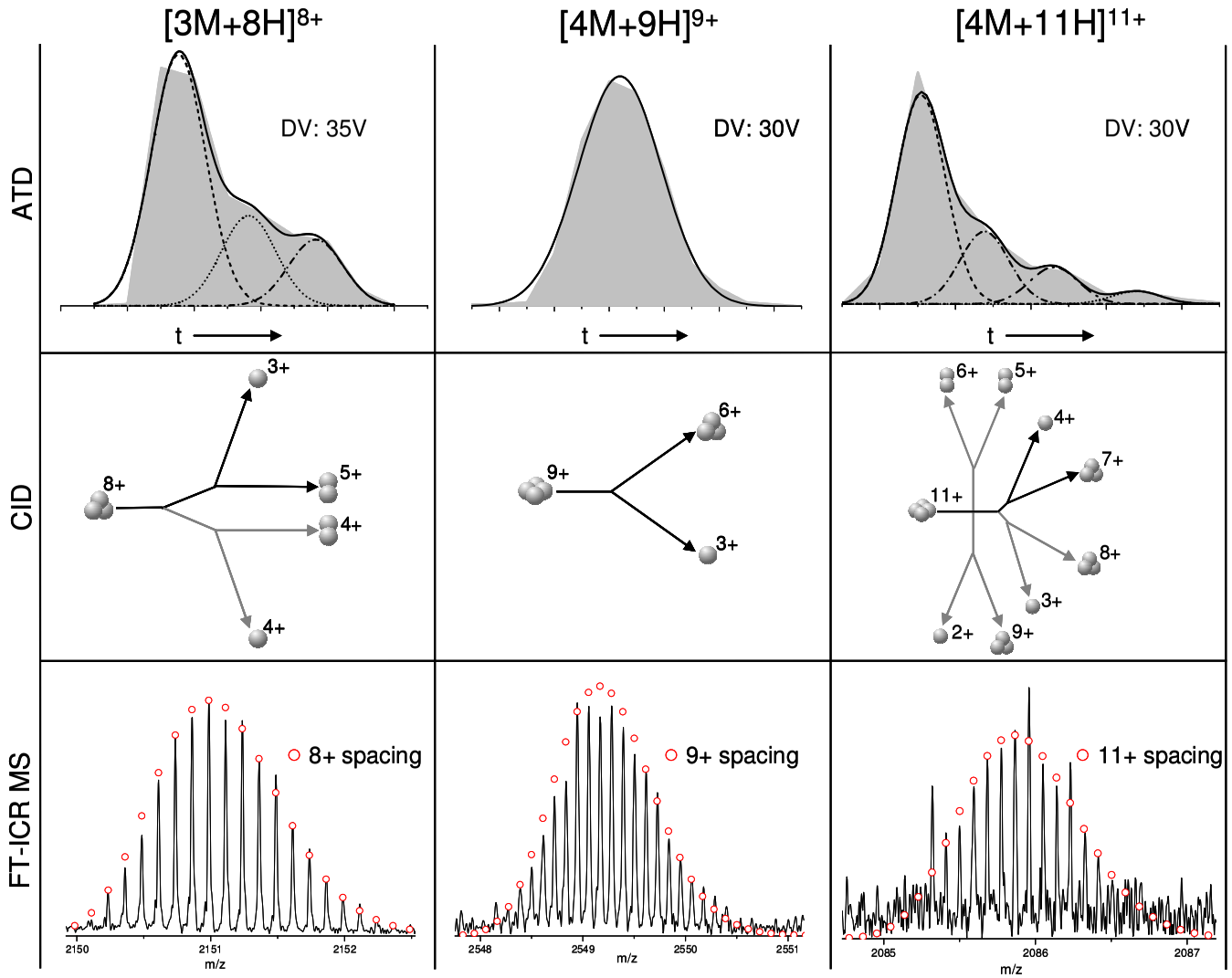


Figure 4.19 | ATD, FT-ICR MS and CID data for $[3M+8H]^{8+}$, $[4M+9H]^{9+}$ and $[4M+11H]^{11+}$.

For ATD data the experimental data are the shaded grey area. Dotted and dashed lines represent the Gaussian fits for each species. The solid black line is the cumulative fit of the Gaussian curves.

For CID data black lines denote the dominant fragmentation pathway.

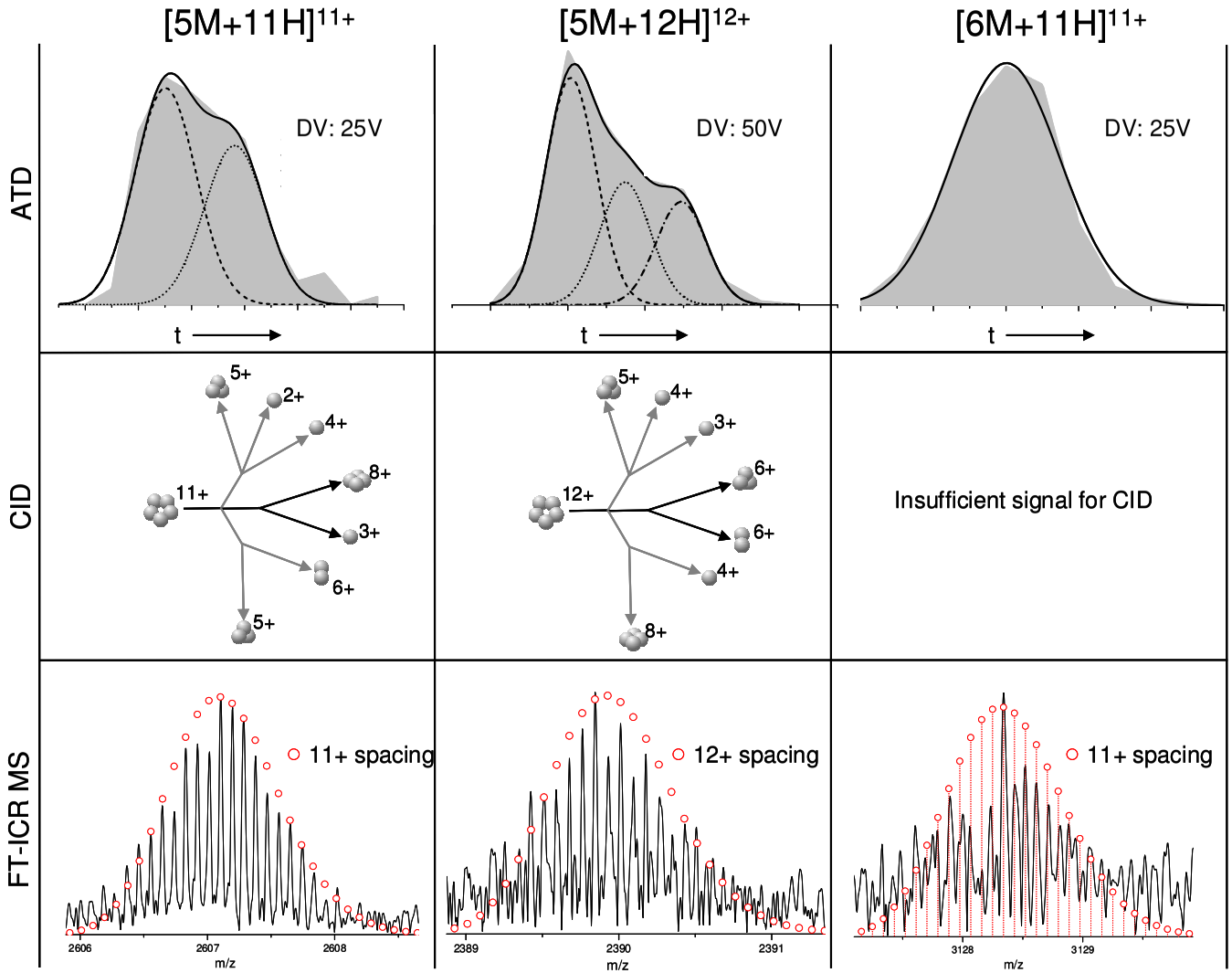


Figure 4.20 | ATD, FT-ICR MS and CID data for $[5M+11H]^{11+}$, $[5M+12H]^{12+}$ and $[6M+11H]^{11+}$.

For ATD data the experimental data are the shaded grey area. Dotted and dashed lines represent the Gaussian fits for each species. The solid black line is the cumulative fit of the Gaussian curves.

For CID data black lines denote the dominant fragmentation pathway.

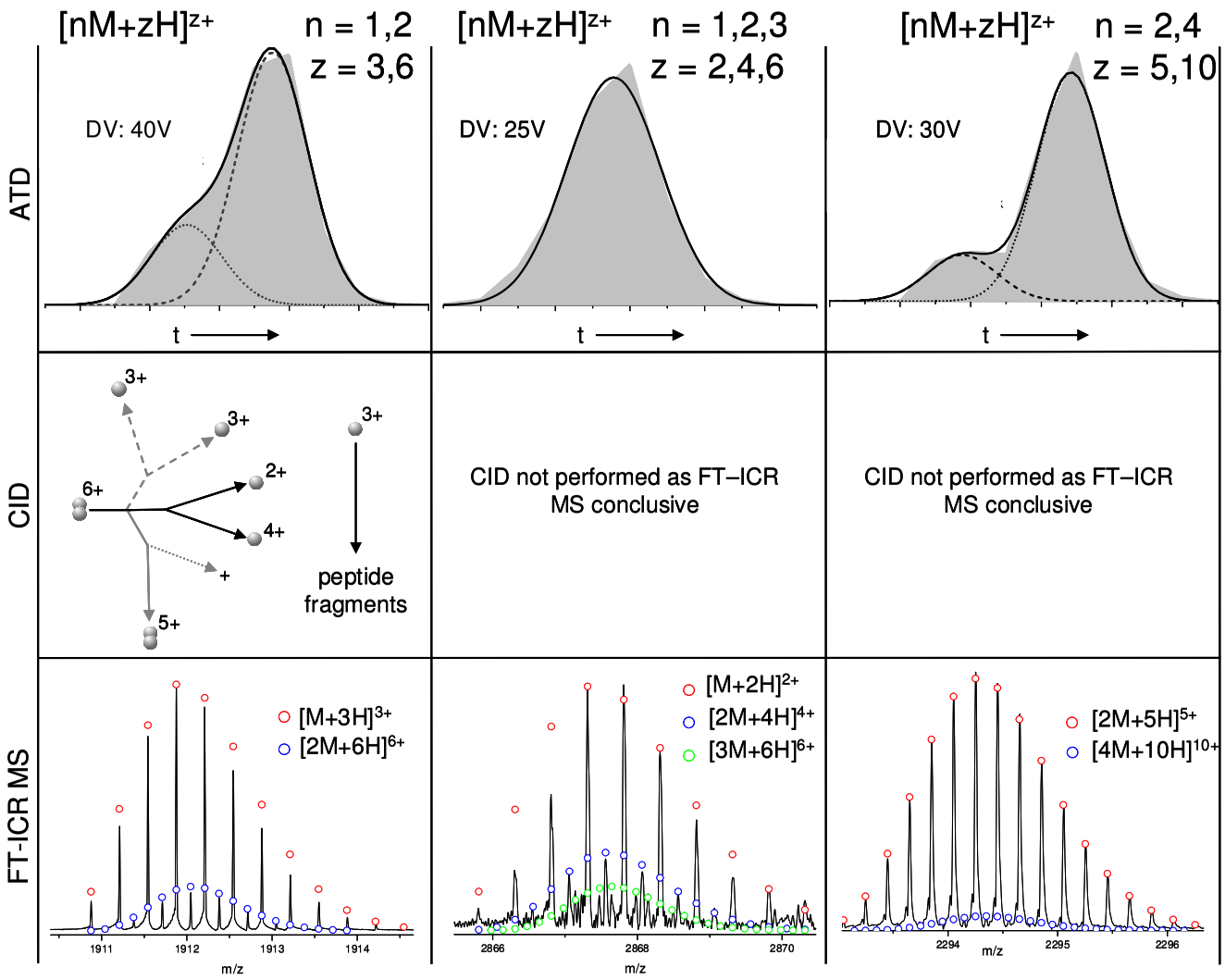


Figure 4.21 | ATD, FT-ICR MS and CID data for $[nM+zH]^{z+}$ where n = 1, 2 and n = 3, 6; n=1, 2, 3 and z =2, 4, 6 and n = 2, 4 and z = 5, 10.

For ATD data the experimental data are the shaded grey area. Dotted and dashed lines represent the Gaussian fits for each species. The solid black line is the cumulative fit of the Gaussian curves.

For CID data black lines denote the dominant fragmentation pathway, dotted lines represent species which cannot be observed by MS. Dashed lines represent product ions which cannot be distinguished from parent ions.

4.3.2.2. Collision Induced Dissociation Experiments

CID experiments provide information on parent ion stability; however they can also aid parent ion identification. An exemplar of this is Figure 4.21, where the oligomeric order n = 1 or 2 with associated charge z = 3 or 6. In the mass spectrum the dominant species is $[M+3H]^{3+}$ so when two peaks are observed in the corresponding ATD the question arises whether compact and extended conformations of monomers are present or whether the more mobile species is the *m/z* coincident dimer $[2M+6H]^{6+}$. Mass isolation of this *m/z* peak and fragmentation of the corresponding species reveals $[2M+5H]^{5+}$ and $[M+4H]^{4+}$ product ions. Neither of these could have been produced by the dissociation of $[M+3H]^{3+}$, confirming the presence of a coincident *m/z* dimer - $[2M+6H]^{6+}$.

Conversely the ATD of $[2M+7H]^{7+}$ (Figure 4.18) exhibits three distinguishable species, one of which could be assigned to coincident oligomer $[4M+14H]^{14+}$, however following CID only monomeric product ions were formed, indicative that the multiple ATD peaks result from distinct conformations of $[2M+7H]^{7+}$. Six other oligomers: $[2M+9H]^{9+}$, $[3M+8H]^{8+}$, $[3M+10H]^{10+}$, $[4M+11H]^{11+}$, $[5M+11H]^{11+}$ and $[5M+12H]^{12+}$ exhibit multiple peaks in their ATDs which CID results (lack of product ions of higher oligomeric order than the postulated parent ion) suggest are different conformations of one species.

4.3.2.3. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

FT-ICR MS can be utilised to obtain a high resolution ^{13}C isotopic mass spectrum which can identify very small populations of oligomeric species at a coincident *m/z* to a dominant spectral peak. Figure 4.21 illustrates this as the isotopic pattern of $[3M+6H]^{6+}$ and $[2M+4H]^{4+}$ species can be observed underlying the principal $[M+2H]^{2+}$ species even though the ATD shows no evidence of the presence of

multiple coincident aggromers. Where sufficient signal was obtained to analyse isotopic distributions, oligomeric species $[2M+7H]^{7+}$, $[3M+8H]^{8+}$, $[4M+11H]^{11+}$, $[5M+11H]^{11+}$ and $[5M+12H]^{12+}$ produced isotopic distributions consistent with the elemental composition and charge of that species and that species alone (Figure 4.18, Figure 4.19 and Figure 4.20). By contrast, in Figure 4.21 the species assigned nominally as $[M+3H]^{3+}$ shows an isotopic distribution of a 3+ ion (red circles) and also a contribution from a species with double the charge (blue circles) and hence double the mass - $[2M+6H]^{6+}$. The ATD of this species can therefore be conclusively defined as containing dimer and monomer; the higher charge state of the dimer means that it travels faster through the drift cell.

ATD, FT-ICR MS and CID data are displayed for oligomeric species $[3M+7H]^{7+}$ (Figure 4.18), $[4M+9H]^{9+}$ (Figure 4.19) and $[6M+11H]^{11+}$ (Figure 4.20) which are as the experimentally observed adjacent charge states to species ($[3M+8H]^{8+}$, $[4M+11H]^{11+}$ and $[6M+13H]^{13+}$) which display multiple ATD peaks. These adjacent charge state oligomers possess close to the same number of protonable sites and are thus assumed to be the most structurally similar species to the oligomers which display multiple conformations. At low drift voltages only single conformations are observed, and this is confirmed by FT-ICR MS and CID data.

Spectra from FT-ICR MS are of sufficient resolution to identify whether all the disulphide bridges in each oligomer are intact. For all species present the simulated isotopic trace for the elemental composition of oligomers with all disulphide bridges intact matches the experimental isotopic pattern. Intact disulphide bridges have previously been detected in insulin extracted from insulin fibrils *in vitro* [38] and in *in vivo* amyloid deposits [15]. The presence of these disulphide bonds places important restrictions on the conformational space each oligomer can explore, and indeed these constraints have been observed to limit the configurations of protofilament sheet packing [39]. Therefore these are important restraints and are utilised in the following molecular modelling.

FT-ICR MS and CID experiments, combined with data from ATDs has allowed us to conclude that additional ATD peaks in insulin oligomeric species $[2M+7H]^{7+}$, $[2M+9H]^{9+}$, $[3M+8H]^{8+}$, $[3M+10H]^{10+}$, $[4M+11H]^{11+}$, $[5M+11H]^{11+}$ and $[5M+12H]^{12+}$ should be assigned as extended conformations, as the data show no evidence for aggromer presence. Multiple conformations of an oligomeric species have been reported in small (~6 amino acids), potentially amyloidogenic, peptides [40]. Our experimental timescale for IM-MS measurements is milliseconds and therefore these conformational families are stable over at least that time window.

The difference in the ratio of the intensities of the assigned *m/z* coincident species (for example, in Figure 4.21, the intensity of $[4M+10H]^{10+}$ differs between the ATD and FT-ICR MS data) can be attributed to differences in the transfer optics of the two instruments (the IM-MS instrument possessing a source that has been optimised for the preservation of non-covalent interactions) and the differing timescales for analysis; milliseconds for IM-MS compared to seconds for FT-ICR MS.

A number of previous studies [41,42] have indicated multiple peaks in ATDs of protein oligomers, however the use of CID and FT-ICR MS allow assignments to be based on solid experimental evidence. This is conclusive evidence for multiple, distinct, conformations of oligomeric species whose assignment has been achieved by further experiments. In every case the most compact conformation of the oligomer is present in the highest abundance, with two, three or four other larger conformers present. It is worth noting that the monomeric species only show single distinguishable conformers in each charge state.

4.3.2.4. Coulombic Forces on Oligomers

A large number of charges will increase the CCS of an aggregate due to Coulombic repulsion [43,44], therefore the presence of extended species in the ATDs could be attributed to charge driven structural unfolding in the gas-phase. De la Mora derived an empirical relation for the maximum number of charges on the surface of a protein

which retains its native fold [45]. In Equation 4.1 z is the number of charges and M_R is the mass of the protein (or in this case, the oligomer).

$$z = 0.0778\sqrt{M_R} \qquad \textit{Equation 4.1}$$

Figure 4.22 shows this relation plotted against the experimentally observed charge states for different oligomeric orders.

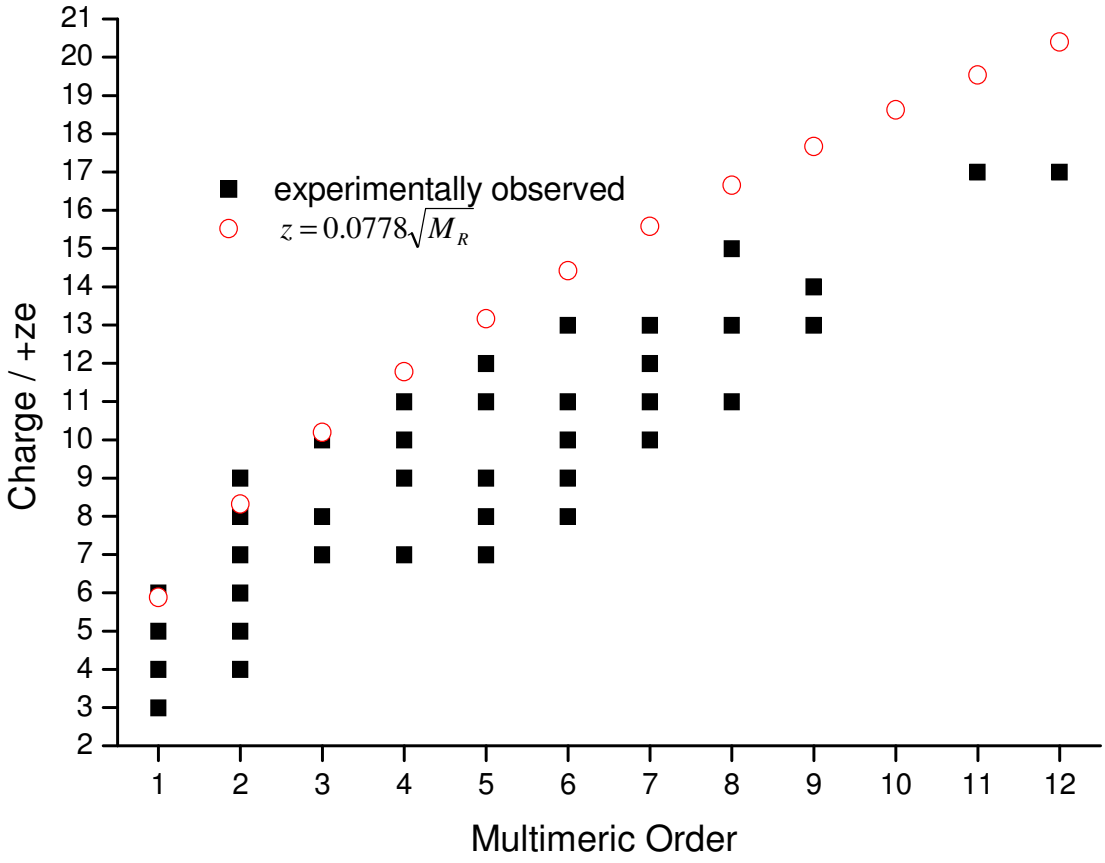


Figure 4.22 | The observed charge carried by each oligomer compared to the de la Mora relation for the maximum number of charges able to be carried on the surface of a protein which retains its native fold.

With the exceptions of $[M+6H]^{6+}$ and $[2M+9H]^{9+}$ the observed charge on all insulin oligomer species is below the limit which can be accommodated on the surface of a protein of this size without Coulombic unfolding. Therefore the range of observed

conformers for each oligomer, and the differences in the charge oligomers carry, are likely to represent inherent structural variations present in solution rather than structures generated by Coulombically driven gas-phase unfolding. This is supported by several studies that show insulin can adopt multiple conformations at low pH and high concentrations [46,33] and allows us to link the conformations we observe *via* IM-MS with atomistically resolved candidate geometries from MD simulations in implicit solvent.

4.3.2.5. Oligomer Distributions

Cumulative ATDs of each ion present in the mass spectrum were recorded after the same number of scans. These ATDs were fitted with a Gaussian distribution (or distributions if conformers or coincident *m/z* oligomers were present) and the area of the distributions noted. Plotting the areas of ATD distributions *versus* the multimeric order enables the consideration of the relative abundances of each oligomer.

From Figure 4.23 it is observed that the monomer is the most abundant species; presenting in dominant charge states of 3+ and 4+. The dimer is the next most prevalent species. The hexamer appears as an extremely stable species, an observation which is consistent with the storage of insulin as a hexamer *in vivo.* There appears to be a slight preference for trimeric, hexameric, nonameric and dodecameric species compared to the neighbouring oligomeric states, possibly suggesting an increased stability of this trimeric, or multiple of trimeric, arrangement. The tetramer and octamer contribute only weakly to the spectrum, pointing to instability in these species relative to the other oligomeric orders.

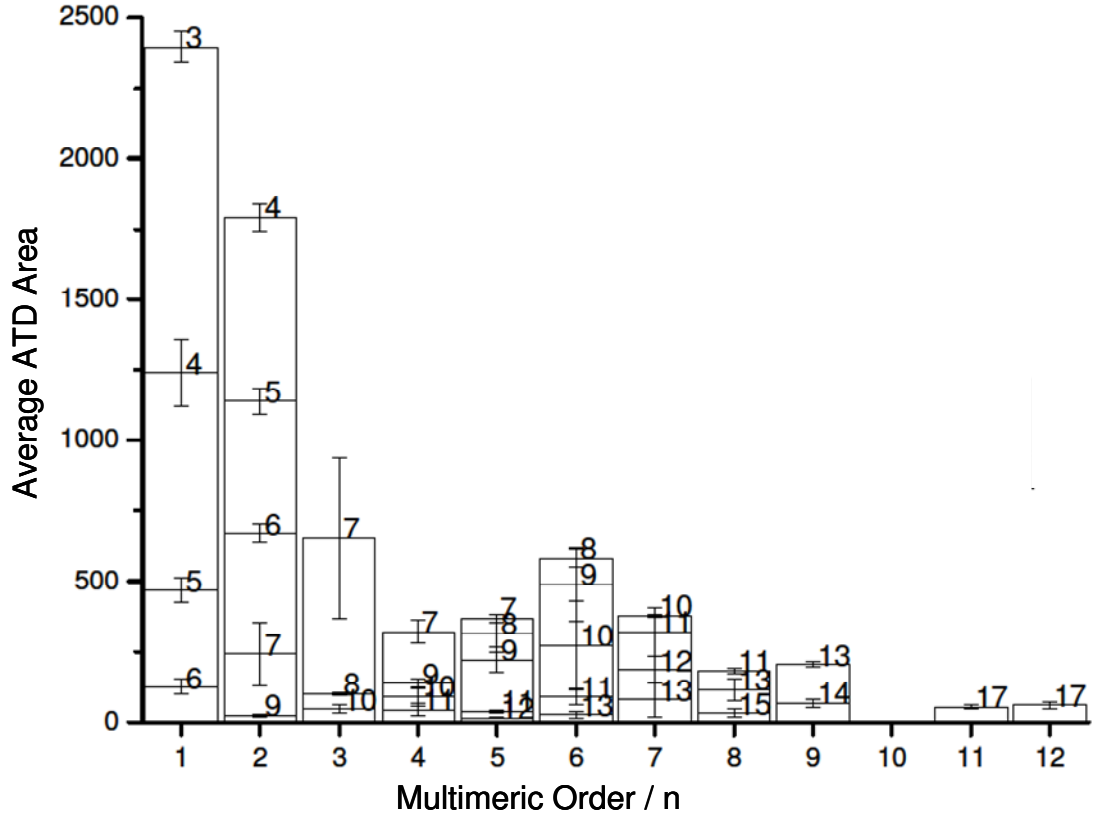


Figure 4.23 | Distribution of Oligomers at t = 0: Each bar represents the integrated area of the ATD for each oligomer at a drift voltage of 50 V. Where more than one charge state is observed the bar is divided proportionately. Positive charge states are indicated by the number above the bar. The data are averaged over three repeats; error bars represent the standard deviation.

Figure 4.23 also shows that some species have an unequal charge state distribution. For the dimeric species $[2M+zH]^{z+}$ is observed with z = 4, 5, 6, 7 and 9. Due to the abundance of the $[2M+7H]^{7+}$ species, the $[2M+8H]^{8+}$ ion is presumed to be present, but is irresolvable in ATD space from $[M+4H]^{4+}$, which is one of the most dominant spectral peaks.

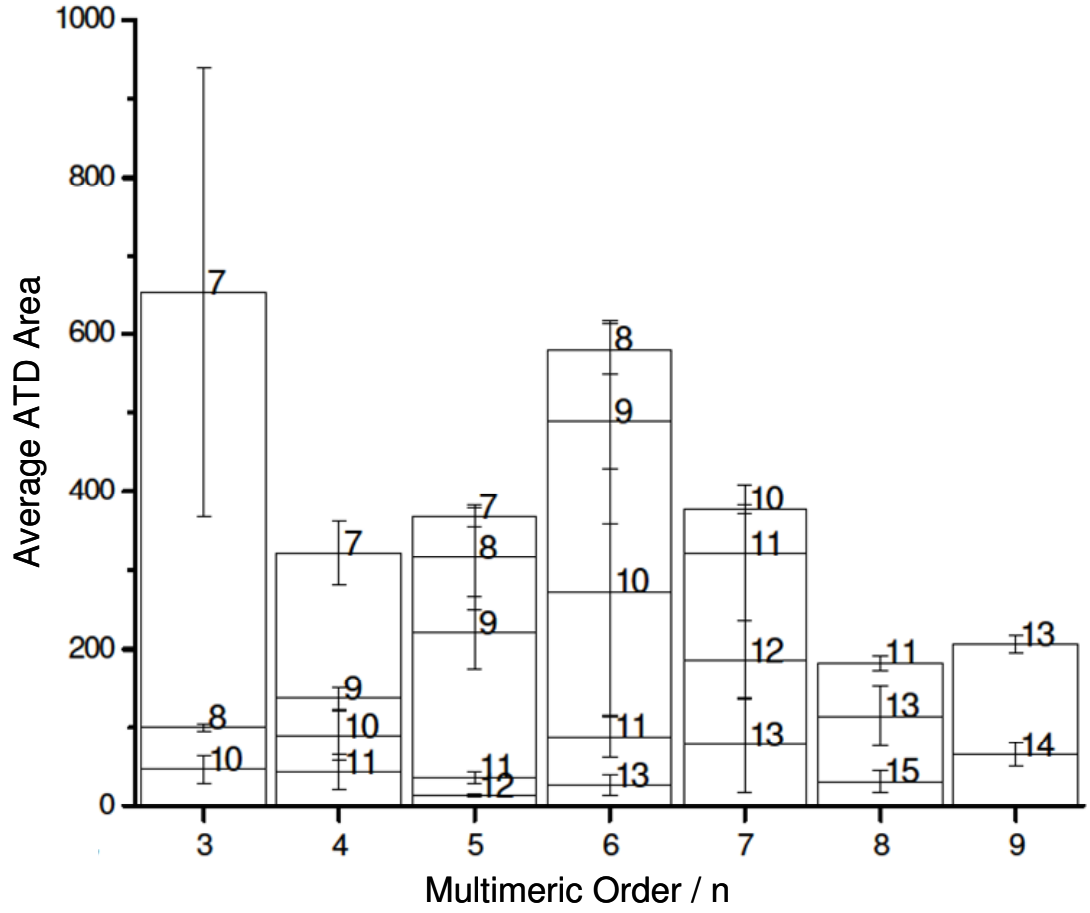


Figure 4.24 | Distribution of Oligomers at t = 0, showing an expansion of the region for n > 2. Each bar represents the integrated area of the ATD for each oligomer at a drift voltage of 50 V. Where more than one charge state is observed the bar is divided proportionately. Positive charge states are indicated by the number above the bar. The data are averaged over three repeats; error bars represent the standard deviation.

4.3.2.6. Collision Cross Sectional Values

We have obtained CCS values for the oligomeric species observed, from the four charge states seen for the monomeric form of insulin to the single charge state of the dodecamer. These are tabulated in Appendix 4 and displayed graphically in Figure 4.25.

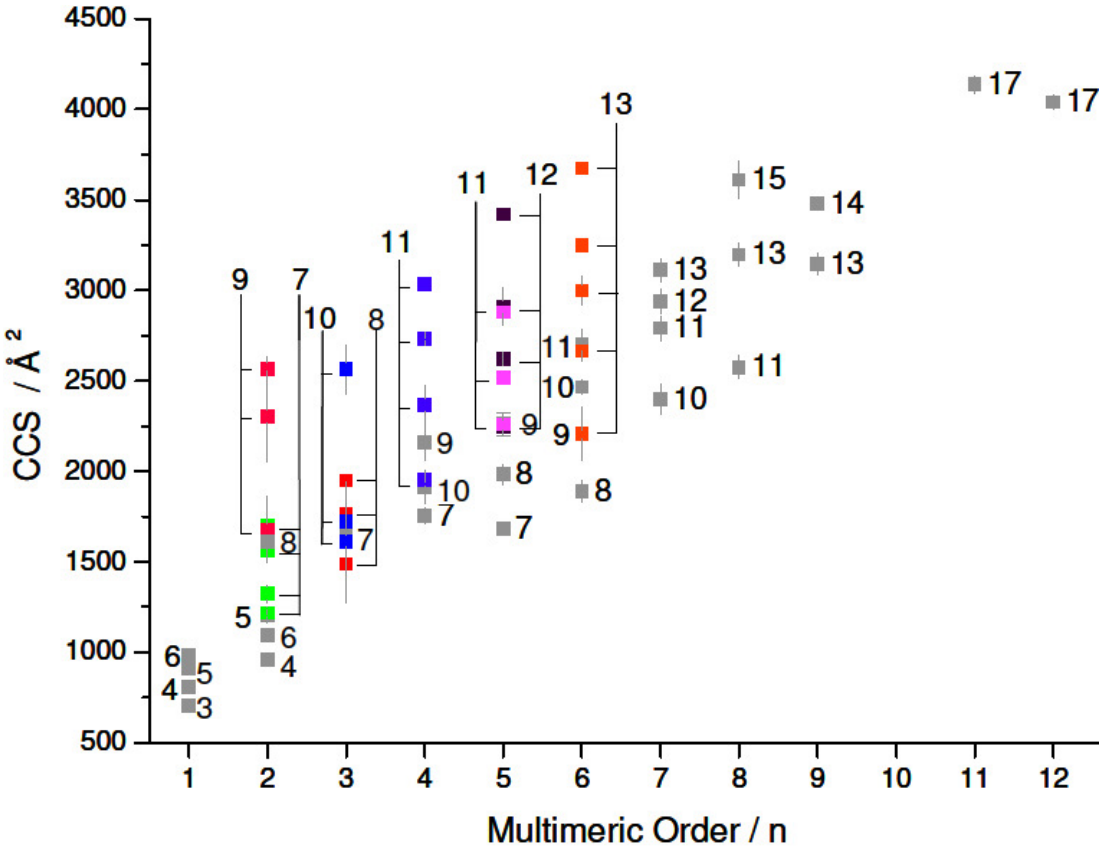


Figure 4.25 | CCSs of oligomers observed at t = 0 for 523 µM insulin in aqueous pH 2 formic acid. Labels are the associated positive charge state of each oligomer. Error bars represent the standard deviation of three repeated measurements. Multiple conformations of the same oligomeric species are coloured: green for $[2M+7H]^{7+}$*, magenta* $[2M+9H]^{9+}$*, red* $[3M+8H]^{8+}$*, blue* $[3M+10H]^{10+}$*, violet* $[4M+11H]^{11+}$*, pink* $[5M+11H]^{11+}$*, purple* $[5M+12H]^{12+}$ *and orange* $[6M+13H]^{13+}$*.*

As observed in the previous chapter for TTR(105-115), an increase in CCS with multimeric order is observed. All species carry more charges than the number of proteins in the oligomer but, as discussed previously, the charges carried by each oligomer are below the limit which would force structural unfolding. Therefore more extended oligomers present in solution, which have a greater number of ionisable sites, acquire more charges in the n-ESI process and thus a correlation between oligomer size and charge state is observed.

However, there are exceptions to this, as the number of solvent accessible sites on a solution structure influences the charge on the gas-phase oligomer. Thus it is possible for the $[2M+5H]^{5+}$ species to have the same CCS as the smallest conformation of $[2M+7H]^{7+}$ due to a lower availability of protonation sites on the former structure. Likewise the CCS of $[5M+9H]^{9+}$ is very similar to that of the smallest conformers of $[5M+11H]^{11+}$ and $[5M+12H]^{12+}$. As $[2M+8H]^{8+}$ is substantially larger than all but the least populated conformation of $[2M+7H]^{7+}$, the presence of an additional protonation site hints that a major structural reorganisation has occurred. $[3M+10H]^{10+}$ is more compact than $[3M+7H]^{7+}$ and some conformers of $[3M+8H]^{8+}$, suggesting a condensed structure which possesses many accessible protonation sites. A similar story occurs for $[4M+10H]^{10+}$ and the smallest $[4M+11H]^{11+}$ conformer which display more compact CCSs than $[4M+9H]^{9+}$. Compact structures which possesses greater availability of ionisable residues, relative to $[6M+9H]^{9+}$ $[6M+10H]^{10+}$ and $[6M+11H]^{11+}$, are observed for the two smallest $[6M+13H]^{13+}$ conformers.

A plateau in CCS for oligomers of the same charge ($[7M+13H]^{13+}$, $[8M+13H]^{13+}$ and $[9M+13H]^{13+}$) is observed, which suggests dense aggregate packing as there is no net increase in the number of protonable sites exposed with oligomer addition. Unexpected decreases in CCSs are observed between $[4M+7H]^{7+}$ and $[5M+7H]^{7+}$, $[5M+8H]^{8+}$ and $[6M+8H]^{8+}$, $[6M+10H]^{10+}$ and $[7M+10H]^{10+}$, $[7M+11H]^{11+}$ and $[8M+11H]^{11+}$ and finally $[8M+13H]^{13+}$ and $[9M+13H]^{13+}$. In some cases the decrease is covered by the standard deviation in both measurements; however the reduction in CCS between $[4M+7H]^{7+}$ and $[5M+7H]^{7+}$, and $[7M+11H]^{11+}$ and $[8M+11H]^{11+}$ is

not, suggesting a more efficient packing and possible structural reorganisation of $[5M+7H]^{7+}$ and $[8M+11H]^{11+}$ compared to their neighbouring lower order oligomers of the same charge.

Oligomers exhibiting multiple conformations can span a large CCS range. The $[2M+7H]^{7+}$ conformational families extend from 1217 to 1701 $Å^2$ and $[2M+9H]^{9+}$ conformations have CCSs between 1679 and 2565 $Å^2$. However, as the charge on $[2M+9H]^{9+}$ is above the de la Mora limit, it is not possible to extrapolate from these gas-phase conformers to solution behaviour. The CCS distribution occupied by oligomers from monomeric to hexameric order is displayed graphically below.

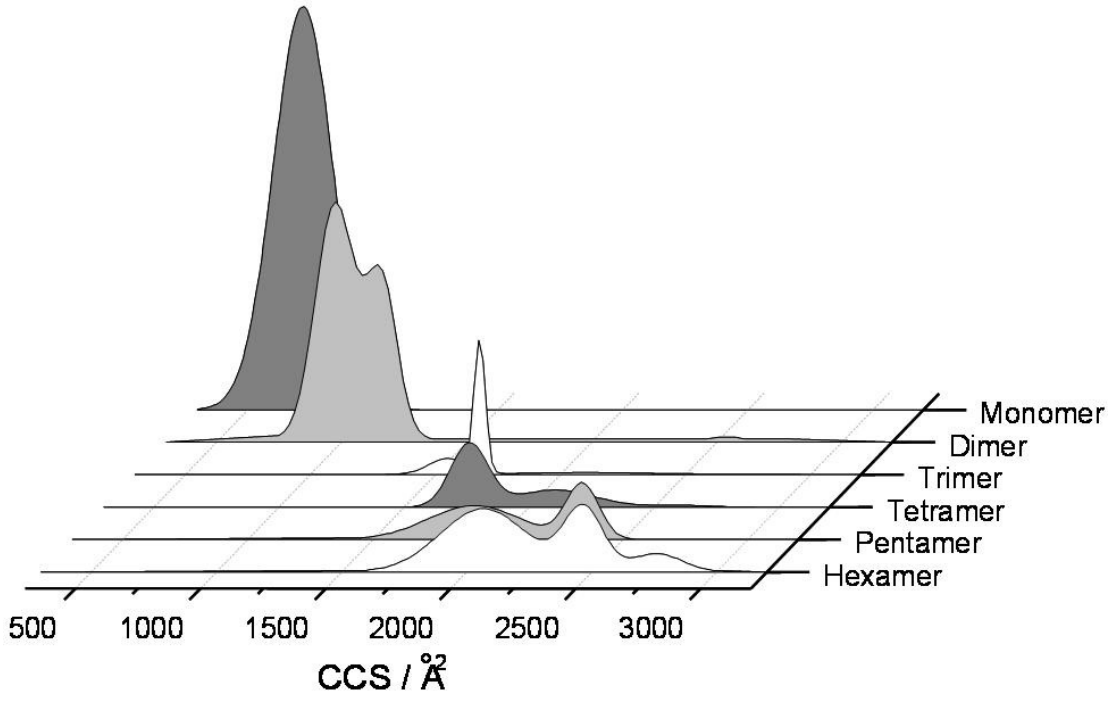


Figure 4.26 | Conformational populations of insulin monomer and each of the oligomers as measured using IM-MS. Data have been obtained from the areas of the deconvoluted peaks in the m/z selected ion ATDs. To aid interpretation, information from each charge state has been merged via a Gaussian fit to provide this composite plot, which highlights the separation capabilities of IM-MS.

Figure 4.26 further demonstrates the separation power of IM-MS, enabling determination of the size distribution of all aggregates (i.e. monomer ions span CCSs of 704 to 978 $Å^2$) but also unequivocally (coupled with high resolution FT-ICR MS) assigns sizes to oligomeric orders (for example the ion at *m/z* 1639 can be identified uniquely as a dimer of CCS 1217 $Å^2$). These data can be compared advantageously to size exclusion chromatography (SEC). Fink *et al.* [47] used SEC to study insulin aggregation, assigning distinguishable 'sizes' to the hexamer, to compact and expanded dimers, and to compact, expanded, and unfolded monomer. With no mass information such data could be misinterpreted: Figure 4.26 demonstrates that conformational space occupied by the monomer and dimer populations significantly overlap, as does the conformational space occupied by species including the trimer, tetramer, pentamer and hexamer.

4.3.2.7. Injection Energy Experiments

Injection energy studies were carried out to ensure that the extended oligomer conformers observed were not caused by structural unfolding in the gas-phase, due to activation of the ions by high-energy collisions as they are transported in the mass spectrometer. IM-MS data were acquired whilst varying the voltage that injects the ions into the drift tube, known as the injection energy. The ATDs of $[2M+7H]^{7+}$, at a drift voltage of 20V, are shown at a range of injection energies in Figure 4.27.

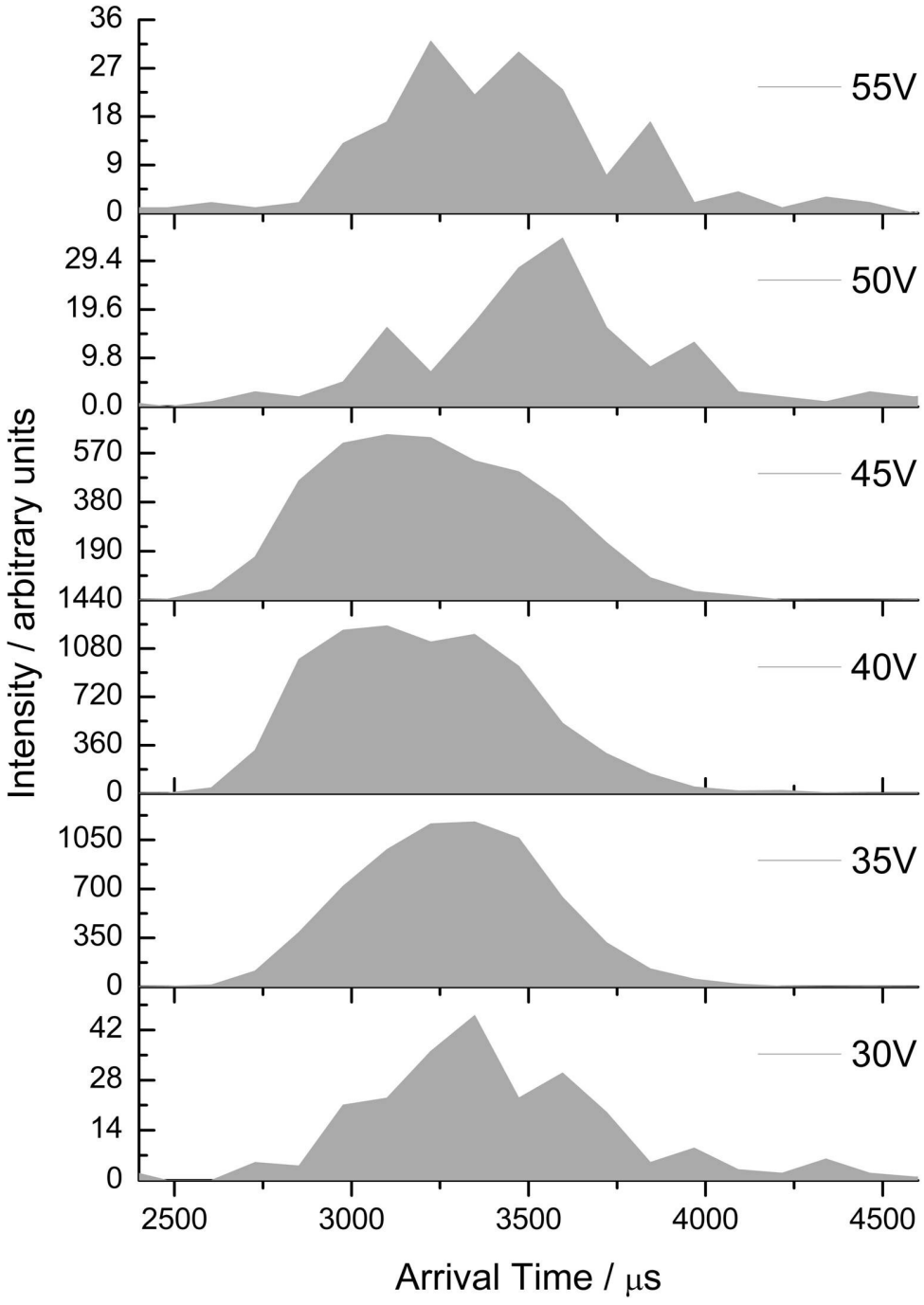


Figure 4.27 | $[2M+7H]^{7+}$ ATDs acquired at a range of injection energies.

The $[2M+7H]^{7+}$ ATDs are shifted to higher arrival times only when the injection voltage is higher than 45 V. This behaviour is replicated in all ATDs which contain

features that FT-ICR MS and CID have identified as extended conformations. As IM-MS experiments were performed with an injection energy of 39.7 V, this indicates that these extended conformers are present in solution, and not a result of structural unfolding in the mass spectrometer.

4.3.2.8. ATD Observations

All ATDs for species of $n > 6$ contain a lower populated peak which arrives before the main species peak (Figure 4.28 and Figure 4.29). This peak is also present in two pentameric and three hexameric species. Due to the high *m/z* values of these species and their relatively high molecular weights, dissection of peak identity by either CID or FT-ICR MS has not been successful. There are two possibilities for the identity of these earlier arriving species; either they are significantly more compact and less populated conformations of the nominal species, or coincident *m/z* higher order aggromers. Both possibilities are shown graphically and discussed below.

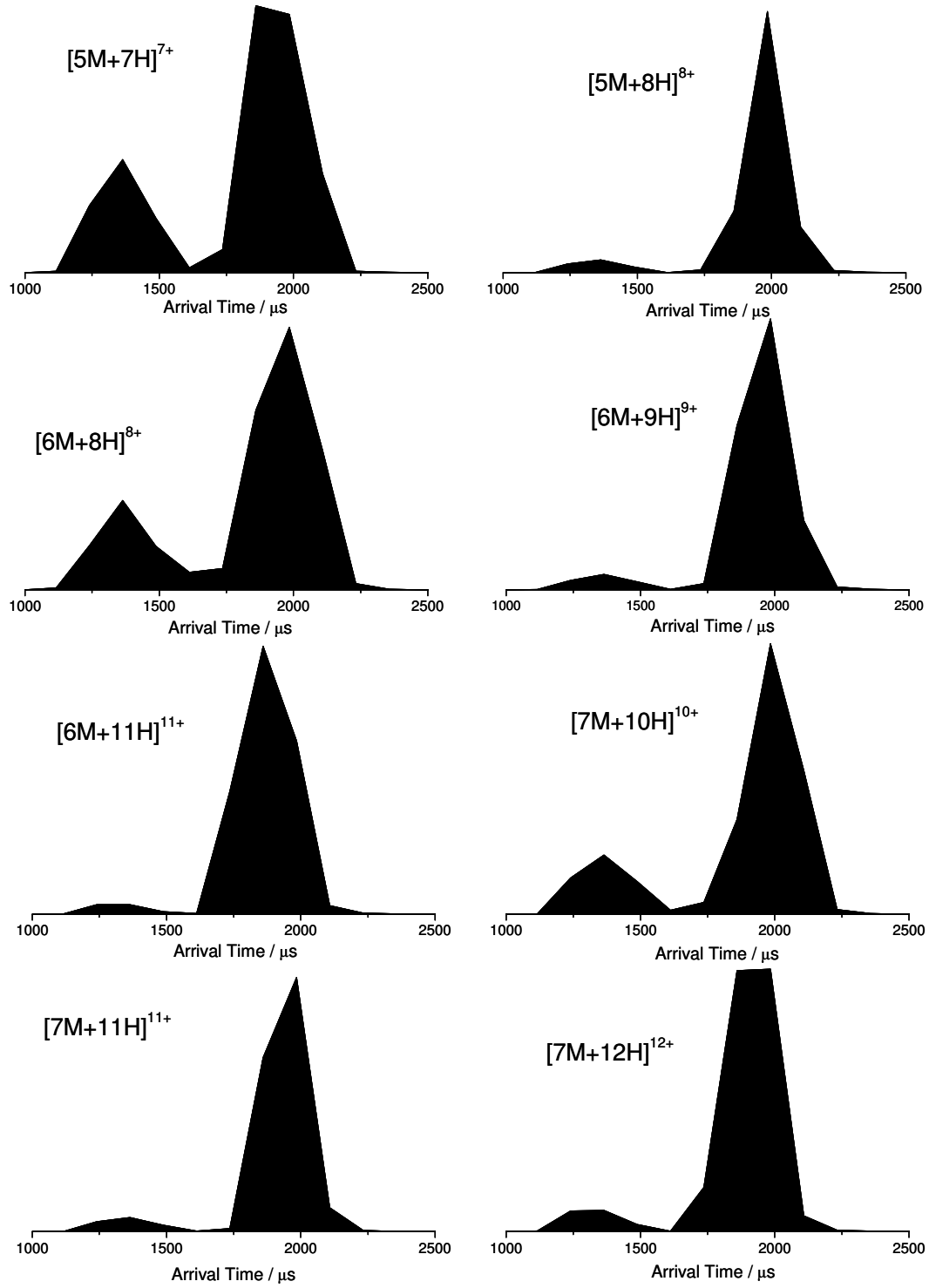


Figure 4.28 | ATDs at a DV of 60 V showing the earlier arriving species present. Labels correspond to the assignment of the later arriving more populated species.

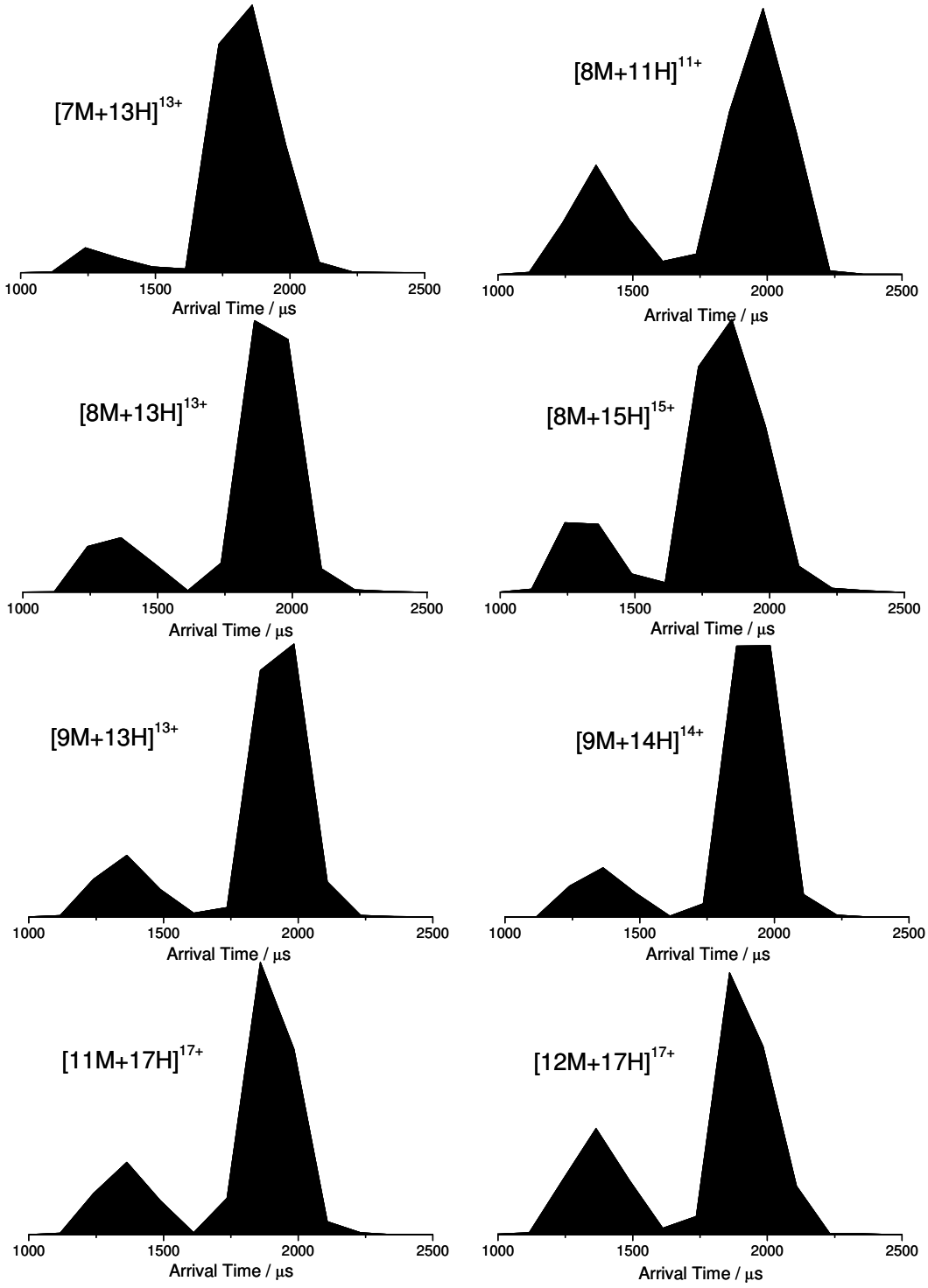


Figure 4.29 | ATDs at a DV of 60 V showing the earlier arriving species present. Labels correspond to the assignment of the later arriving more populated species.

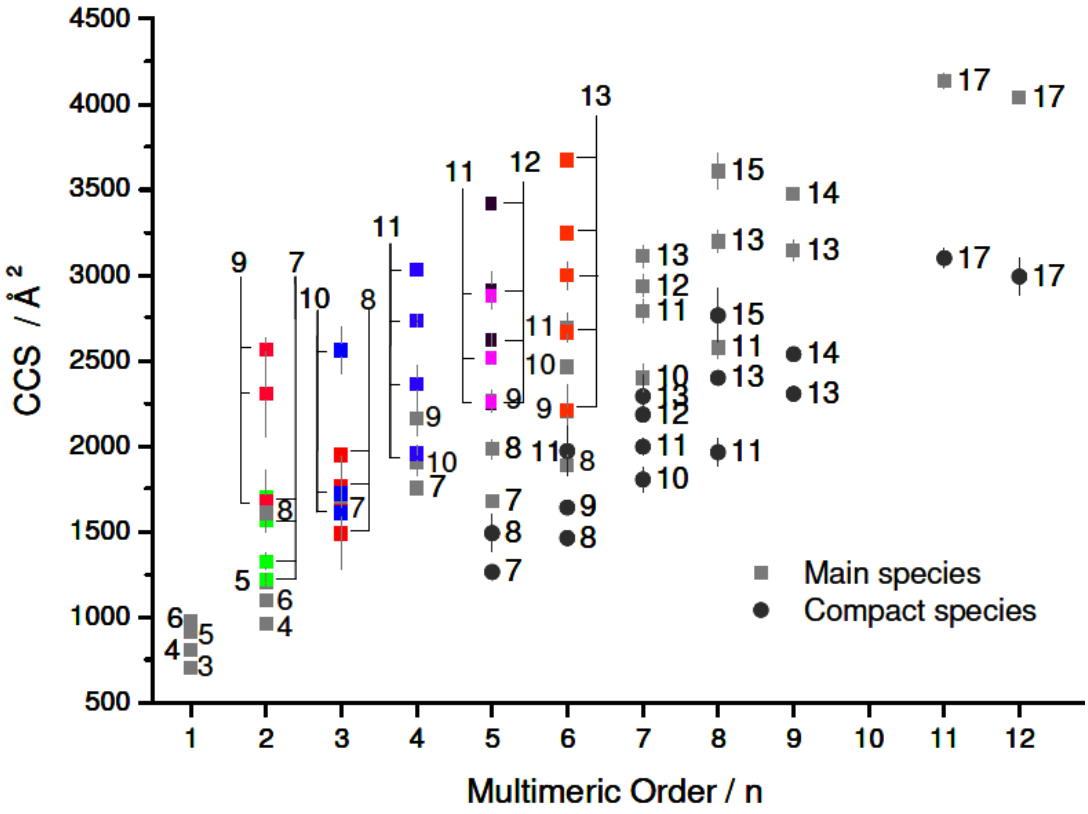


Figure 4.30 | CCSs of oligomers observed at t = 0 for 523 μM insulin in aqueous pH 2 formic acid, with earlier arriving ATD peaks assigned as compact species. Labels are the associated positive charge state of species. Error bars represent the standard deviation of three repeated measurements.

The degree of compaction experienced by the oligomers, for the above assignment to be correct, would be extreme (Figure 4.30). The smallest conformation of $[2M+7H]^{7+}$ and the compact $[5M+7H]^{7+}$ species have CCSs that are the same within experimental error. Hence, using this interpretation of the data, a two and a half times increase in mass would not result in any increase in size. Perhaps more striking is that the compact $[6M+8H]^{8+}$ species possess a CCS substantially smaller than that of the $[2M+8H]^{8+}$. If the oligomers are assumed to be spherical, their density can be calculated. $[2M+8H]^{8+}$ has a density of 0.24 Da/Å^3 whilst a compact $[6M+8H]^{8+}$ species would have a density of 0.82 Da/Å^3, almost three and a half times greater. As average density of globular protein standards has been reported as 0.44 Da/Å^3 [42], a density of 0.82 Da/Å^3 seems unlikely. Thus the assignment of the earlier arriving

ATD peaks as coincident *m/z* higher order oligomers appears more probable, and the resulting CCSs are depicted in Figure 4.31.

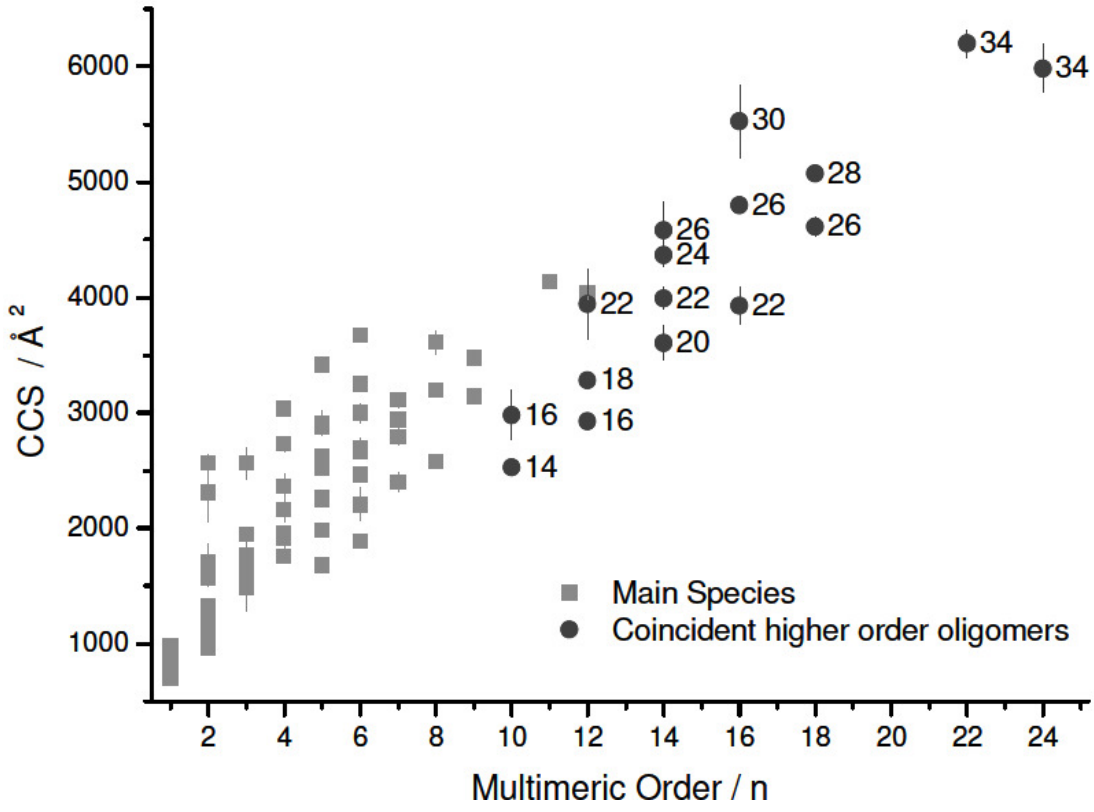


Figure 4.31 | CCSs of oligomers observed at t = 0 for 523 µM insulin in aqueous pH 2 formic acid with earlier arriving ATD peaks assigned as higher m/z species. Labels are the associated positive charge state of species. Error bars represent the standard deviation of three repeated measurements.

The largest oligomer observed would therefore be a 24mer with a molecular weight of 137.6 kDa, to date the largest oligomeric species observed for an amyloidogenic system by IM-MS. The average of the densities of the coincident *m/z* higher order oligomers is 0.43 Da/Å^3, in good agreement with the literature [42], which lends support to this assignment.

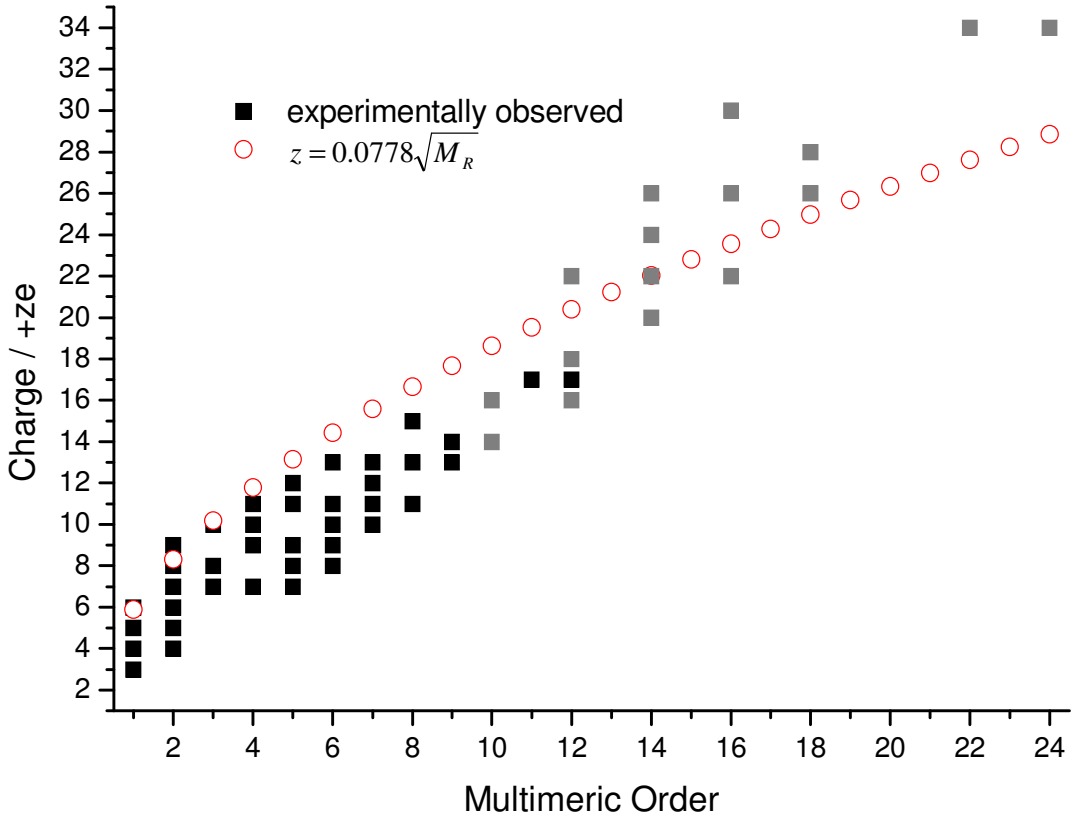


Figure 4.32 | The observed charge carried by each oligomer compared to the de la Mora relation for the maximum number of charges able to be carried on the surface of a protein which retains its native fold. The grey squares are the earlier arriving ATD peaks assigned as higher order oligomers.

The de la Mora relation plotted in Figure 4.32 shows that if the assignment of the earlier arriving ATD peaks as larger oligomers is correct only the $[10M+14H]^{14+}$, $[10M+16H]^{16+}$, $[12M+16H]^{16+}$, $[12M+18H]^{18+}$, $[14M+20H]^{20+}$, $[14M+22H]^{22+}$ and $[16M+22H]^{22+}$ species have a sufficiently low charge to prevent significant Coulombically driven unfolding. Thus differences in the charge these oligomers carry are likely to be due to inherent structural variations present in solution. Of these species; $[10M+16H]^{16+}$ and $[12M+16H]^{16+}$ and $[14M+22H]^{22+}$ and $[16M+22H]^{22+}$ have CCSs which agree with their standard deviations. All other higher order oligomers carry more charges, and thus their CCS is likely to be governed, in part, by

charge driven unfolding. Regardless of this, the CCSs of these higher order oligomers suggest substantial compaction. The $[12M+16H]^{16+}$, $[12M+18H]^{18+}$ and $[12M+22H]^{22+}$ dodecameric species have CCSs which are less than or equal to that of $[12M+17H]^{17+}$, indicating the former possess a greater availability of solvent accessible ionisable residues but have undergone a structural compaction. Other oligomers of higher than dodecameric order display the expected correlation between increasing charge and increasing CCS.

4.3.3. Molecular Modelling

4.3.3.1. MD of Oligomers from Monomeric to Hexameric Order

The initial coordinates files used to carry out MD simulations were those of the crystal structure of dimeric bovine insulin, pdb code 2ZP6 [4]. Starting with this structure several multimeric species were constructed, from monomer up to hexamer, by applying crystallographic symmetry transformations given in the pdb file. After removing the zinc di-cations and submitting the structures to the web site H++ [48] (imposing a pH of 2), all the structures were minimised *in vacuo*. Related CCS values were calculated using MOBCAL software [49], implementing the trajectory method (TM). The results are given in Table 4.1.

Multimeric order	monomer	dimer	trimer	tetramer	pentamer	hexamer
						
CCS ($Å^2$)	815.89	1318.6	1729.5	2119.1	2428.7	2586.7

Table 4.1 | CCSs for the oligomeric species from monomeric to hexameric order. A structure surface representation for all the species is displayed: the monomeric units closer to the reader are depicted in orange, further away units are depicted in green.

Following minimisation, the dimer has a calculated CCS of 1319 $Å^2$, very similar to that of one of the experimentally measured $[2M+7H]^{7+}$ conformations (1324 $Å^2$). Thus the second most populated gas-phase conformation of $[2M+7H]^{7+}$ occupies a volume similar to that of the crystal structure.

4.3.3.2. Monomeric Species $[M+3H]^{3+}$ and $[M+4H]^{4+}$

The total charge of dimeric bovine insulin in solution at pH 2 is +8 if all of the acidic side chains are protonated. However the insulin dimer extracted from the hexameric crystal structure pdb 2ZP6 [4] yielded a net charge of +7 on submission to the webserver H++ [48], as Glu4 from the A chain of one of the monomers is surrounded by many positively charged side chains, and thus remains unprotonated. This is in support of the high population of $[2M+7H]^{7+}$ observed by MS and IM-MS. The dimer was split into two monomeric species $[M+3H]^{3+}$ and $[M+4H]^{4+}$, facilitating further investigation of the very compact (1217 $Å^2$) and highly extended (1709 $Å^2$) conformations adopted by $[2M+7H]^{7+}$. $[M+3H]^{3+}$ and $[M+4H]^{4+}$ were subjected to MD for 152 ns using the Amber10 software package [50], implementing the Amber ff99SB-ILDN force field [51]. During dynamics, snapshots were collected every 2 ps to give an ensemble of 76,000 structures for each monomeric species.

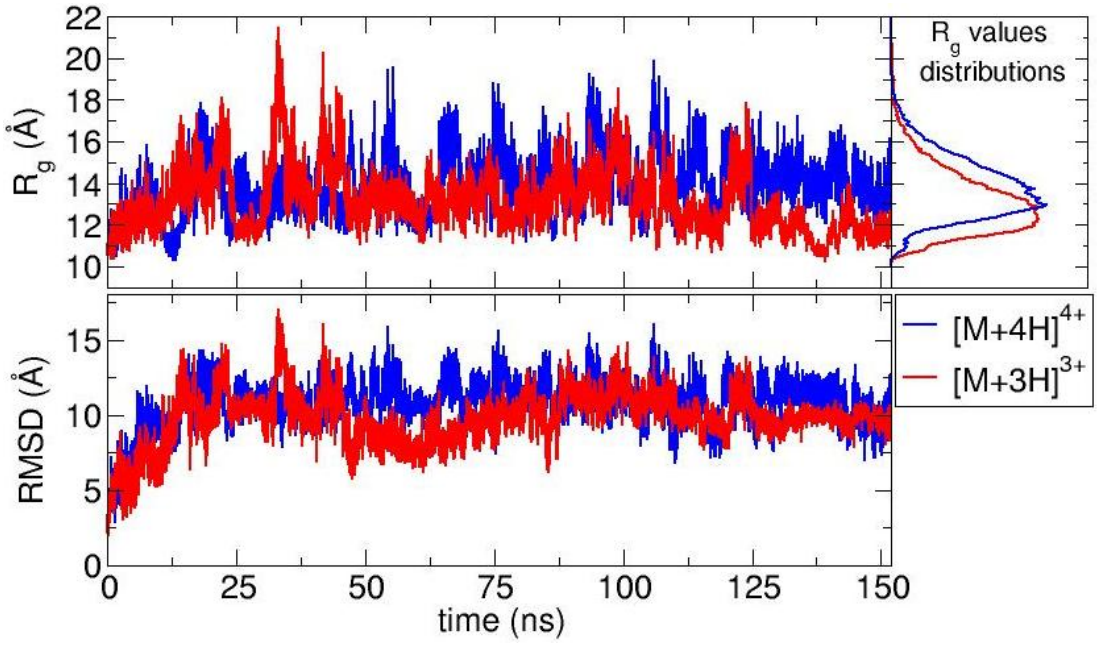


Figure 4.33 | R_g and RMSD time series recorded for 150 ns of dynamics for $[M+4H]^{4+}$ and $[M+3H]^{3+}$.

Both monomeric species sampled complex conformational landscapes (Figure 4.33): R_g spanned values between 10 and 22 Å, whereas the backbone root mean square deviations (RMSDs) span values up to 15 Å. As expected the $[M+4H]^{4+}$ species explores, on average, a larger conformational space due to its higher charge.

4.3.3.3. Correlation Between CCS and R_g

Through a simulated annealing dynamics loop, 500 structures of $[M+3H]^{3+}$ were stored. The implemented loop scheme is detailed in Section 4.2.6.2 and the implicit solvent model adopted is described in Section 4.2.6.1 with a 0.002 ps time step. The runs were conducted with NAMD [52] software. For the 500 obtained structures of the $[M+3H]^{3+}$ monomer, R_g and CCS were calculated. The related scatter plot is given in Figure 4.34 together with the line of the linear regression (which has a correlation coefficient equal to 0.83).

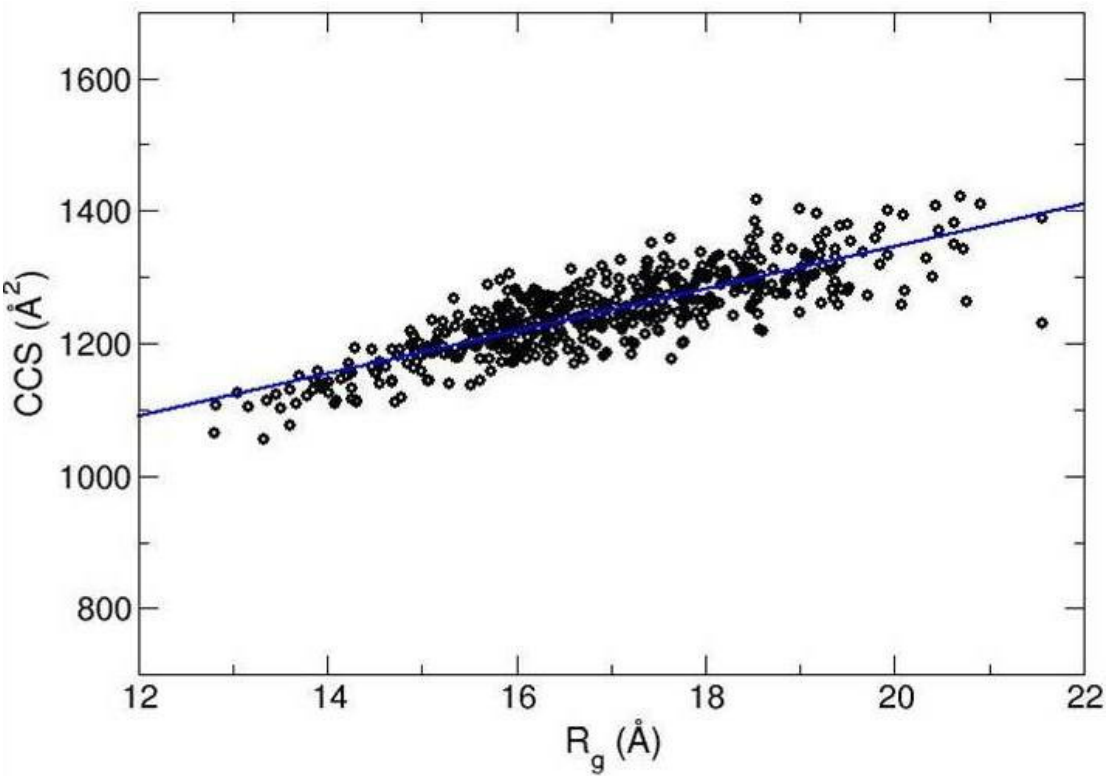


Figure 4.34 | Correlation between R_g and CCS for simulated $[M+3H]^{3+}$ structures. The values were derived from a simulated annealing procedure in implicit solvent.

A linear correlation between R_g and CCS can be observed in Figure 4.34, demonstrating the two variables are comparable.

4.3.3.4. Cluster Analysis

The first 2 ns of the dynamics run were discarded, leaving 75,000 structures of each monomeric species from the original ensemble. A cluster analysis [53] was performed on these structures, which derived 10 clusters of conformations. To construct the similarity matrix of distances the backbone RMSD between the pairs of structures was measured. This procedure resulted in 10 conformational families, whose populations are given in Table 4.2.

Clusters	I	II	III	IV	V	VI	VII	VIII	IX	X
$[M+3H]^{3+}$	27.7	25.1	18.7	7.8	6.8	5.8	3.7	1.8	1.8	0.8
$[M+4H]^{4+}$	23.9	22.0	14.5	8.9	6.5	6.0	6.0	5.9	5.8	0.4

Table 4.2 | Cluster populations (given in %) of conformations sampled along the MD runs of the two monomeric species.

4.3.3.5. Protein-Protein Docking

To elucidate the structure of the most abundant and most compact dimer, $[M+3H]^{3+}$ and $[M+4H]^{4+}$ species with the lowest calculated R_g were used in a protein-protein docking procedure, implementing the webserver Hex [54], based on both shape and electrostatic correlation. After docking, the dimer with the lowest R_g was selected and subjected to dynamics *in vacuo* for 15 ns. The dynamical evolution of its CCS is reported in Figure 4.35. Despite the fact that our simulation time is short compared with the experimental timescale, and too short to sample the conformational change that occurs to a desolvating protein effectively [55], we find conformations which are consistent with the experimental results.

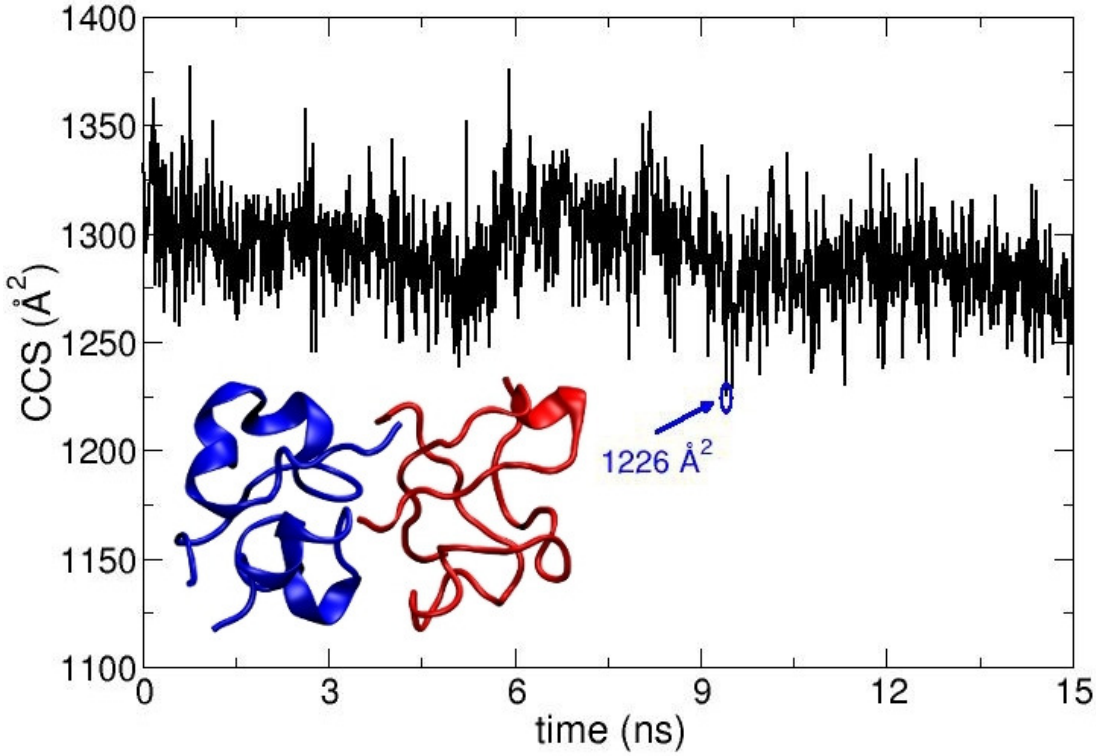


Figure 4.35 | Gas-phase dynamical evolution of the CCS for the most compact dimer. The structure with the lowest CCS is displayed. The experimental CCS for the most dominant, most compact, $[2M+7H]^{7+}$ species is 1217 ± 33 Å².

Across the dynamics trajectory a correlation between secondary structure and compactness exists. This also holds true for the constituent monomers and in Figure 4.36 and Figure 4.37 we show the scatter plot of R_g values versus helicity, beta and turn content, assessed through the DSSP algorithm [56]. For both $[M+3H]^{3+}$ and $[M+4H]^{4+}$ the proportion of β-strand secondary structure is notably higher for the smallest explored conformations, especially for the $[M+3H]^{3+}$ species. Quantitatively, if R_g values are divided into quartiles, from the most compact to the most extended structures, the average percentages of β-strand content are 1.74, 0.88, 0.55 and 0.21 for $[M+4H]^{4+}$ and 14.09, 8.81, 5.63 and 3.04 for $[M+3H]^{3+}$. Other types of secondary structure are more equally distributed along the R_g values (Figure 4.36, Figure 4.37). It should be reiterated that the native state of insulin is predominantly α-helical, and that, both experimentally and computationally, all of the disulphide bonds remain intact within the oligomers.

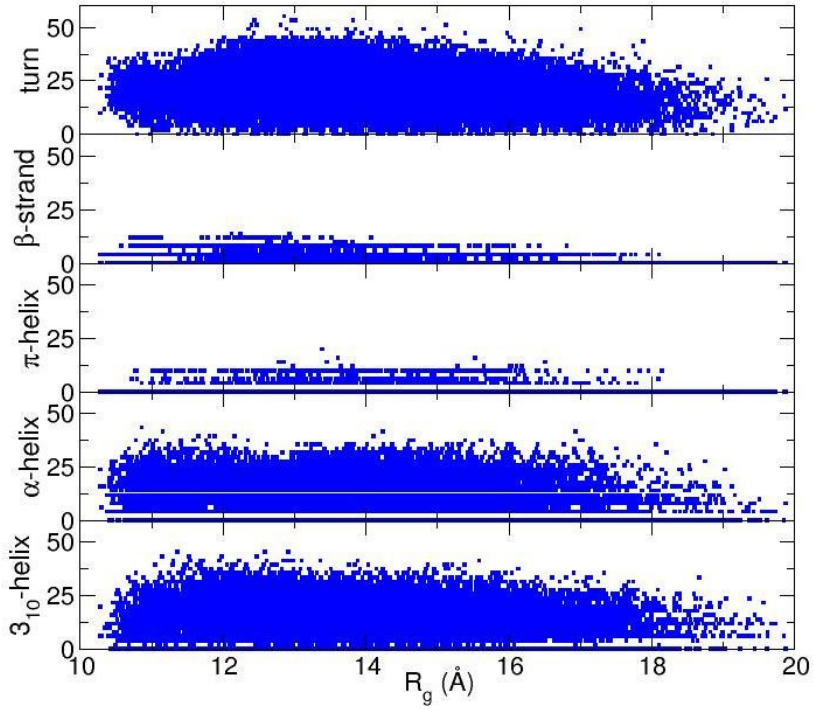


Figure 4.36 | R_g versus secondary structure percentage content for $[M+4H]^{4+}$.

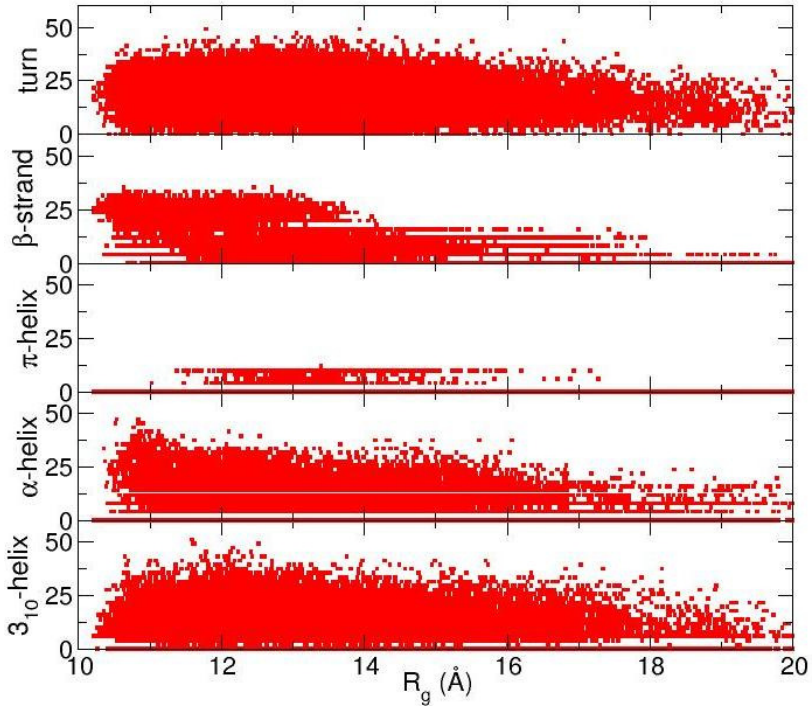


Figure 4.37 | R_g versus secondary structure percentage content for $[M+3H]^{3+}$.

As shown above, by selecting and docking the most compact monomers, MD simulations provide conformations of the insulin dimer that correspond to the compact species we observe experimentally. For all oligomeric species observed experimentally, the most significantly populated conformation is the most compact form, and the above simulations indicate that these may possess enhanced β-sheet structure.

Of the 10 conformational families (I-X) for both $[M+3H]^{3+}$ and $[M+4H]^{4+}$ the largest (I and II) each contain on the order of 25 % of the conformations sampled for each species (Table 4.2). Figure 4.38 displays the representative structures for the most populated cluster of each monomer species.

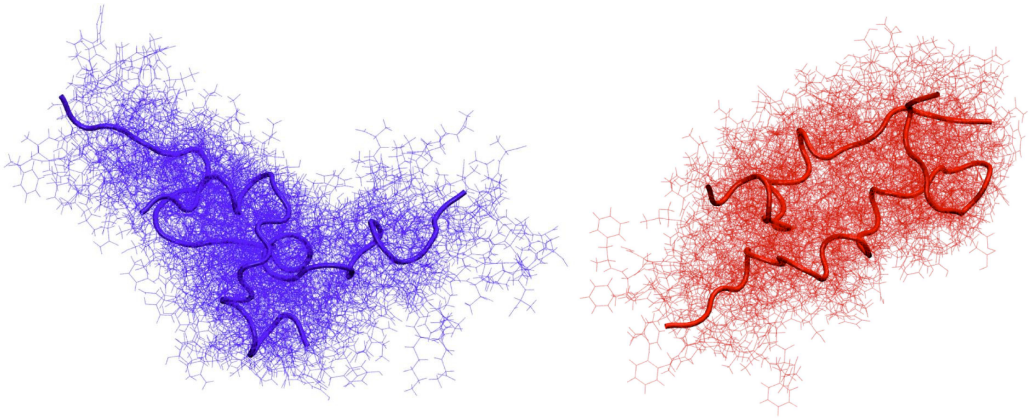


Figure 4.38 | Representative structures (tubes) and conformations (lines) of the most populated cluster for $[M+4H]^{4+}$ (blue) and $[M+3H]^{3+}$ (red).

Representative structures of these most populated clusters were assembled using the same protein-protein docking approach. The resulting 1000 'best docking' dimers were minimised *in vacuo* and the CCSs calculated for the 100 species lowest in energy (Figure 4.39).

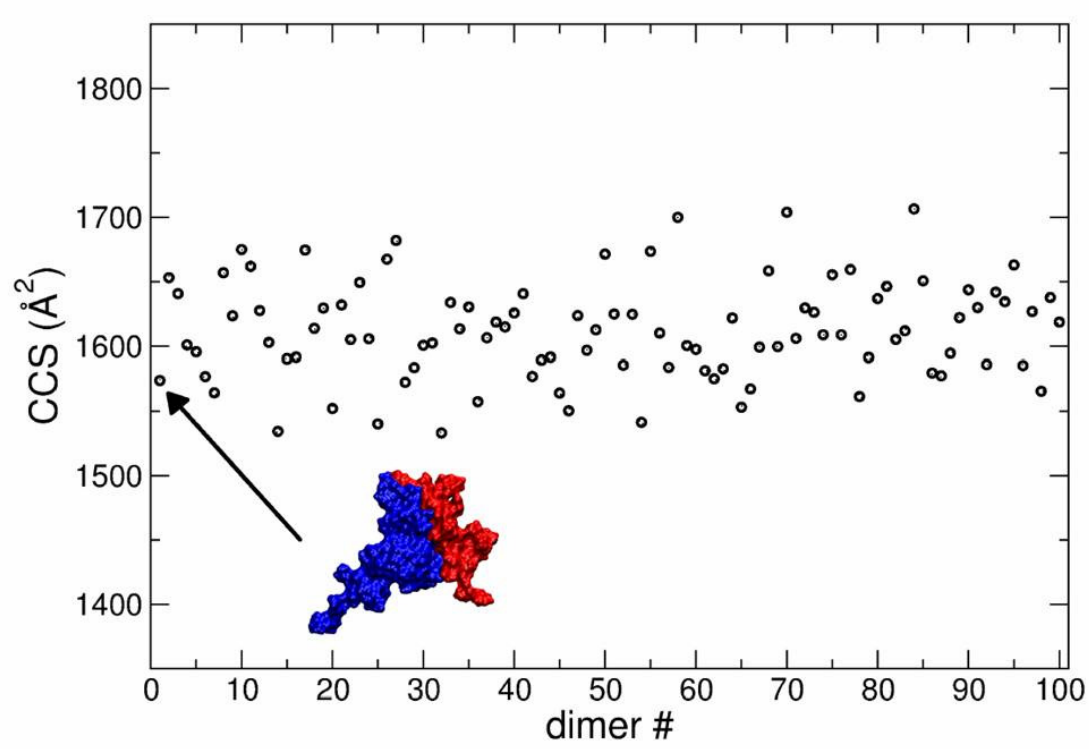


Figure 4.39 | CCS values of the dimers derived from the protein-protein docking between the representative structures of the most populated families. The lowest energy "best docking" dimer is displayed with a surface representation with $[M+4H]^{4+}$ shown in blue and $[M+3H]^{3+}$ in red.

The corresponding values (Figure 4.39) distribute around a mean value of 1612 ± 38 Å^2, in agreement with the experimentally measured CCS of 1564 ± 10 Å^2 for the $[2M+7H]^{7+}$ second largest conformational family.

4.3.3.6. Contact Interface and Stability of Dimers Derived from Docking

To analyse the interacting residues at the interface between assembled monomeric units, the distribution of the α carbon pairwise (CA-CA) distances over the 1000 structures derived from the docking procedure have been calculated. The notation $m_i^{(j)}$ was assigned to the monomeric units, where i = I, II, . . . , X characterises a representative structure of one of the ten clusters reported in Table 4.2 (only the first

two most populated clusters have been considered in this case ie. i = I and II). The notation j represents the considered net charge (i.e. j = 3 or 4). The cut off for the CA-CA distance to be included was set equal to 10 Å.

The representative structure of the most populated cluster was assembled with itself and with the representative one of the second most populated, obtaining six families of 1000 dimers each: $m_I^{(3)}m_I^{(3)}$, $m_I^{(3)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(3)}$, $m_I^{(4)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(4)}$ and $m_I^{(4)}m_{II}^{(4)}$. A dimer is selected from each $m_i^{(j)}m_k^{(l)}$ family which is representative of the most frequent monomer monomer interface and is the lowest in energy. The resulting distributions are projected onto a xy plane, where x and y are the numbered sequences of the first and second monomer respectively. This procedure enables the selection of the lowest energy dimers from the most frequently represented monomer-monomer interface.

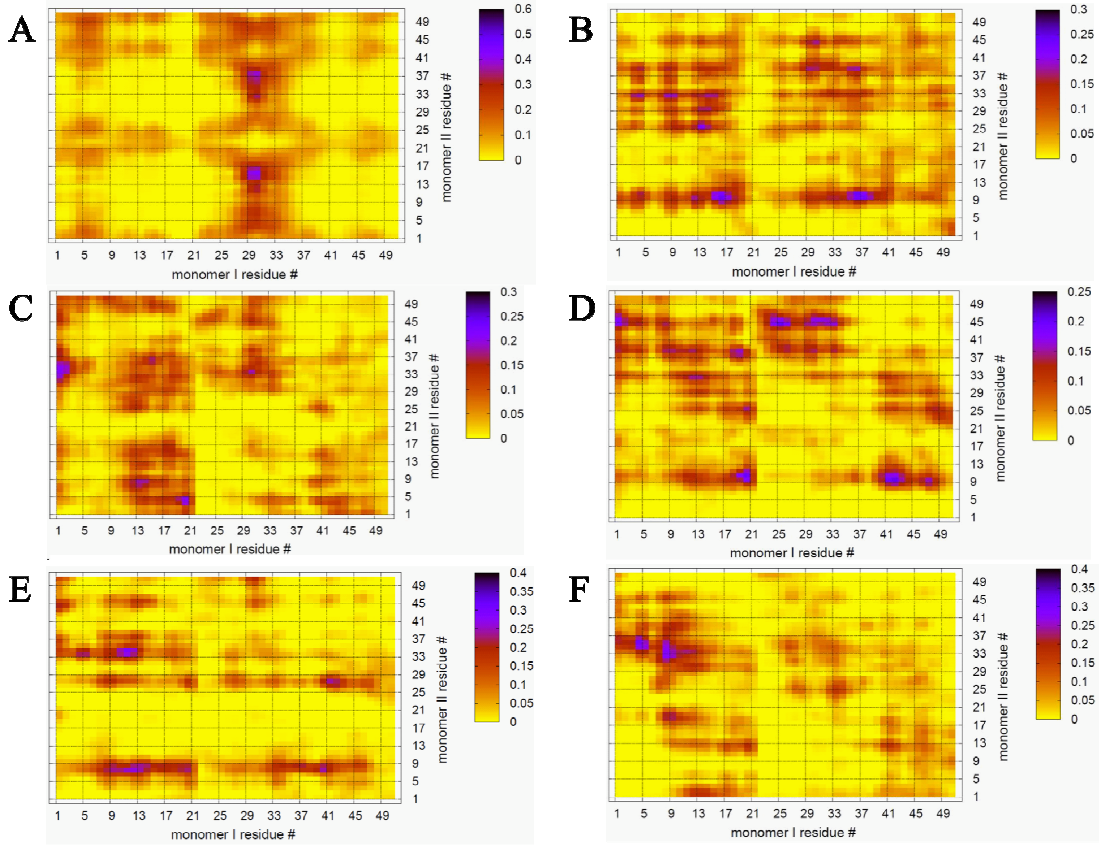


Figure 4.40 | xy projection of the distributions of α carbon pairwise distances. The two upper, middle and lower figures are related to the $[2M+6H]^{6+}$, $[2M+7H]^{7+}$ and $[2M+8H]^{8+}$ species respectively. In particular the figures A, B,

C, D, E and F refer to families of docked monomers named $m_I^{(3)}m_I^{(3)}$, $m_I^{(3)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(3)}$, $m_I^{(4)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(4)}$ and $m_I^{(4)}$ $m_{II}^{(4)}$ respectively.

The stability of each lowest energy dimer was evaluated by testing the interaction energy between the monomers that form dimers with the Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) calculation implemented in the Amber software package [57]. Dimers were immersed in a box of TIP3P [29] water molecules and a neutral charge achieved by the addition of chloride ions, further details are given in Section 4.2.6.3. Using this approach, the binding energy of two monomeric structures that form a dimer can be calculated, overcoming the "static" picture resulting from docking by including dynamical features of the derived dimers.

Even though the energy evaluation is overestimated in absolute value, as the entropic contribution is neglected, this approach gives a sufficiently good representation of the relative trend between similar systems [58], like the selected dimers in this work.

The results of the calculations are shown in Figure 4.41; all of the binding free energies are negative, testifying that the dimers derived from the protein-protein docking are stable. Furthermore, the more compact dimers are the most stable ones with the most negative binding energies. This correlates with our experimental observation of the most compact dimers being the most abundant.

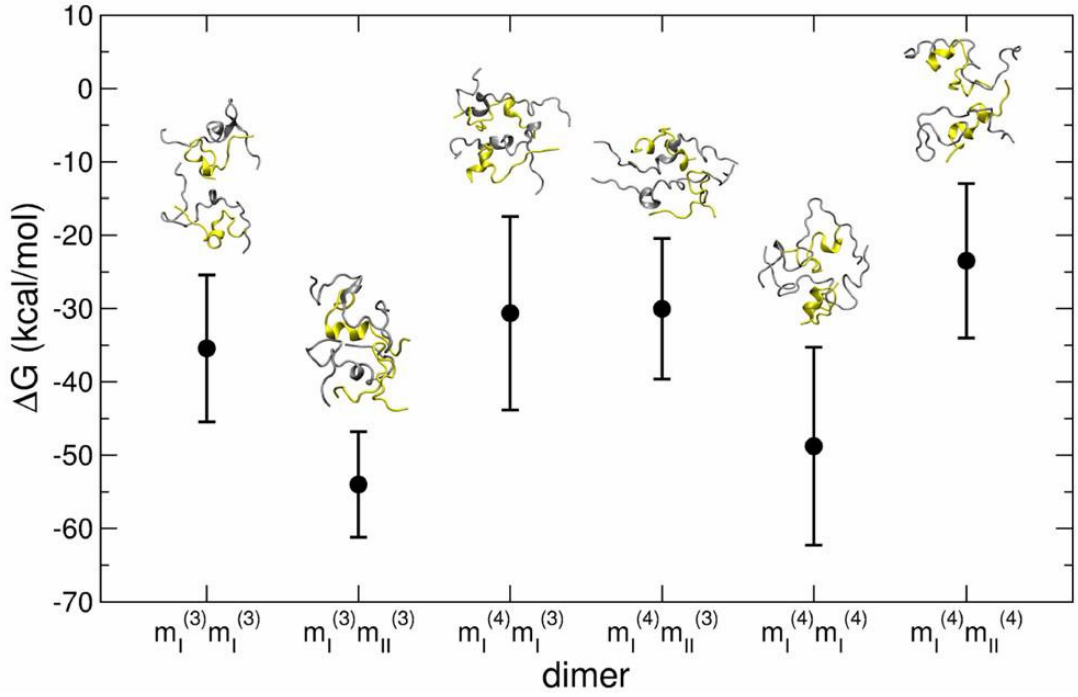


Figure 4.41 | Values of binding energy between monomers forming the selected dimers. The binding energy calculated via MM-PBSA for each selected dimer ($m_i^{(j)}m_i^{(j)}$) is depicted along with the final structure for each from the MD trajectory. A and B chains are displayed in yellow and silver respectively

Analysis of the contributions to the binding energy indicates that in all six cases hydrophobic interactions add favourably to the driving force for self assembly (Table 4.3). This does not hold true for the hydrophilic interactions, which are negative in only two out of six cases. This confirms that hydrophobic interactions are fundamental in stabilising the assembly of the monomeric units. Even though the docking procedure was performed using monomers unfolded in implicit solvent, and hence likely to favour surface-exposed hydrophilic residues and burial of hydrophobic groups, the MM-PBSA procedure has revealed the importance of the hydrophobic interactions in the monomer-monomer contact interface.

	A	B	C	D	E	F
Hydrophobic (kcal/mol)	-6.8	-8.7	-7.4	-6.2	-10.4	-4.9
Hydrophilic (kcal/mol)	103.6	13.0	21.6	-27.7	49.9	-66.9

Table 4.3 | Hydrophobic and hydrophilic contributions to the solvation energy of the representative dimer structures. The values are derived from the Poisson-Boltzmann (PB) and surface area (SA) approaches respectively, with the MM-PBSA procedure. Letters A, B, C, D, E and F refer to the representative dimers of the families $m_I^{(3)}m_I^{(3)}$, $m_I^{(3)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(3)}$, $m_I^{(4)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(4)}$, *and* $m_I^{(4)}$ $m_{II}^{(4)}$ *respectively.*

The interaction interface is retained during each 25 ns simulation, as shown by the residue specific average <RMSD> value (Figure 4.42). The highest values of the <RMSD> are from residues that do not participate in the interface; these are located at the termini of the both chains, although there is greater variability in the B chain (Figure 4.41). The interface between monomers in the different dimers is not identical, with both A-B/A-B and A-B/B-A stacking mechanisms observed.

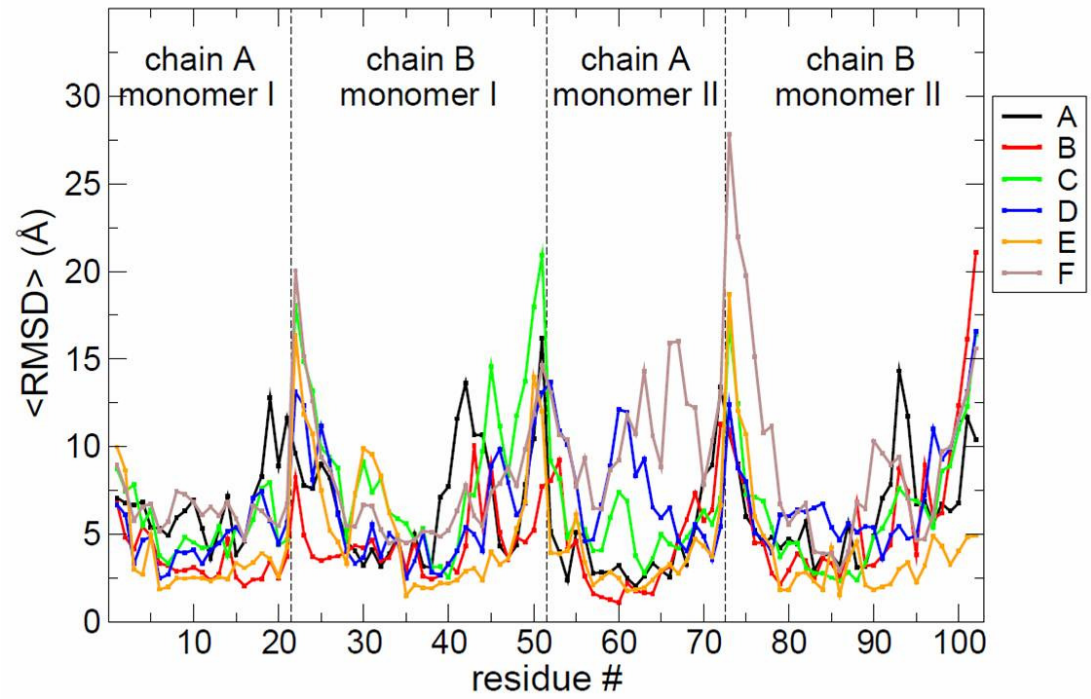


Figure 4.42 | Residue number versus the <RMSD>. Lines marked with letters A, B, C, D, E and F in the legend box refer to selected dimers from families $m_I^{(3)}m_I^{(3)}$, $m_I^{(3)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(3)}$, $m_I^{(4)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(4)}$, *and* $m_I^{(4)}\ m_{II}^{(4)}$ *respectively.*

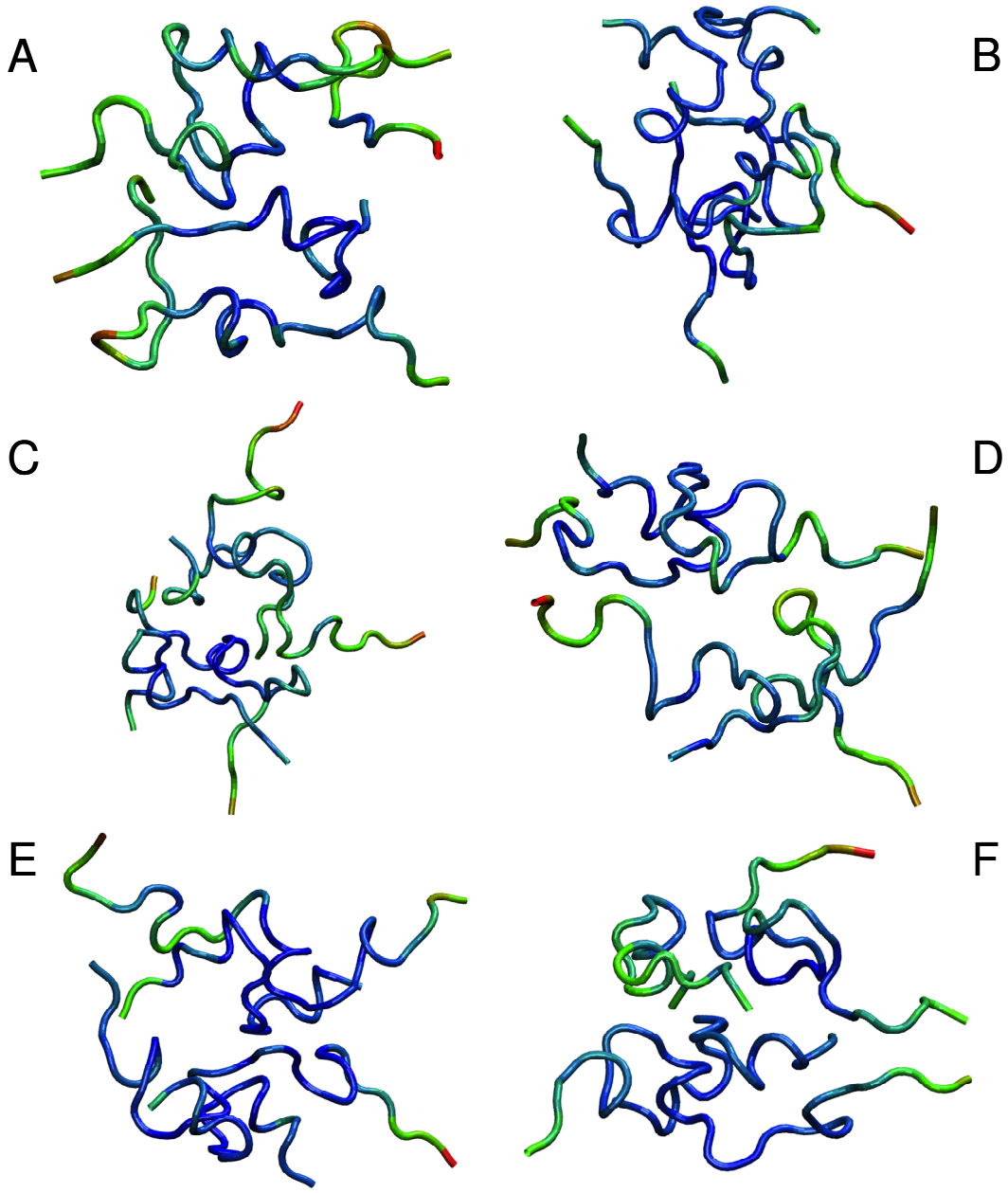


Figure 4.43 | Structures with residues coloured to show the <RMSD> where blue is lowest, green intermediate and red largest. Letters A, B, C, D, E and F in the legends box refer to the selected dimers from the families $m_I^{(3)}m_I^{(3)}$, $m_I^{(3)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(3)}$, $m_I^{(4)}m_{II}^{(3)}$, $m_I^{(4)}m_I^{(4)}$, *and* $m_I^{(4)}\ m_{II}^{(4)}$ *respectively.*

During the dimer dynamics 25 structures were saved (one every nanosecond), from which the water molecules were stripped out. These structures were then minimised *in vacuo* and their CCS calculated. The results are displayed in Figure 4.44, CCS

values between ~1500 and ~1700 Å^2 are obtained, which are in good agreement with experiment.

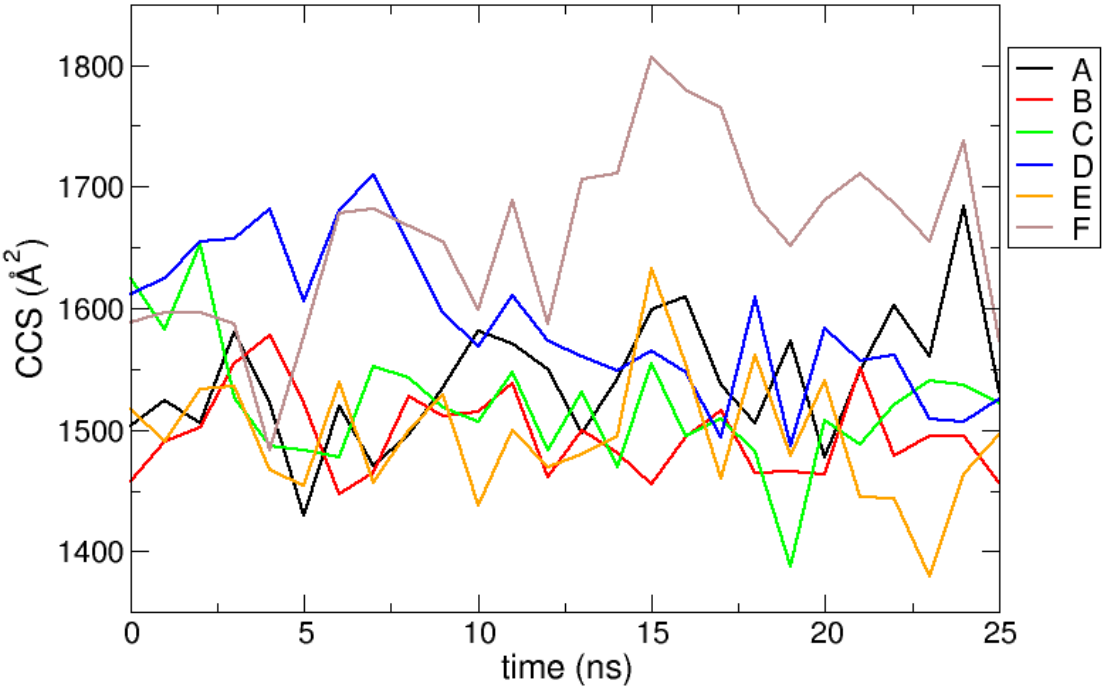


Figure 4.44 | The CCS evolution of selected dimers from families $m_I^{(3)}m_I^{(3)}$(A), $m_I^{(3)}m_{II}^{(3)}$ (B), $m_I^{(4)}m_I^{(3)}$ (C), $m_I^{(4)}m_{II}^{(3)}$ (D) $m_I^{(4)}m_I^{(4)}$ (E) and $m_I^{(4)}m_{II}^{(4)}$ (F).

4.3.3.7. Comparison with CryoElectron Microscopy

Using low resolution 3D structures of fibrils, Jiménez *et al.* [39] postulate that a single protofilament of insulin consists of stacked monomers, and further that each insulin molecule occupies two layers, where the A and B chains of insulin stack on top of the other, connected by disulphide bridges. From the given electron densities of the single protofilament they estimate a dimension of 30 x 40 Å^2 [39]. From this, if we assume that the inter layer spacing is less than the smallest dimension, which it must be if the layers are connected by disulphide bridges, we can obtain a value for R_g of ~19.5 Å for the smallest unit that forms these protofilaments. From the relationship we have derived between R_g and CCS (Figure 4.34), we can compare this to our measured values. The most extended monomer from our implicit solvent calculations

has an R_g of 21.4 Å and a corresponding CCS of ~1391 Å^2. The average values of R_g for monomers are 13.2 Å (CCS 1120 Å^2), and 13.9 Å (CCS 1160 Å^2), for $[M+3H]^{3+}$ and $[M+4H]^{4+}$ respectively (Figure 4.33). Results from IM-MS show that monomeric species never have a CCS that exceeds 1000 Å^2, and the most populated have a lower CCS than this (Appendix 4). The dimeric species we observe experimentally have CCS that are more comparable to the R_g derived from the data of Jiménez et al. [39], and interestingly our most populated gas-phase dimers correspond to R_g between 12.8 and 36.3*. It is tempting to speculate that the structures in the gas-phase where the dielectric constant is 1 may be more similar to those present in the hydrophobic environment of the protofilaments [59], and indeed our data show that a dimeric building block cannot be ruled out. If, as postulated by Jiménez *et al.* [39,] the protofilament is comprised of stacked monomers, these must elongate substantially as the fibrils grow. IM-MS shows that monomers only exist as a single conformational family (Figure 4.26), whereas higher order oligomers have several stable conformers over the millisecond IM-MS experimental timescale, which may reveal the extended forms required for fibril growth.

4.3.3.8. Simulating Extended Dimers

By selecting and docking $[M+3H]^{3+}$ and $[M+4H]^{4+}$ structures with the highest R_g (Figure 4.33, structure at ~33 ns for $[M+3H]^{3+}$ and ~106 ns for $[M+4H]^{4+}$) in an identical fashion to that described previously, dimers with CCSs which reproduce the experimental values of the larger $[2M+7H]^{7+}$ conformations can be simulated. The results are presented in Figure 4.45 and the mean value is equal to 1874.3 Å^2. This average differs by ~170 Å^2 from the target experimental value of 1701.0 Å^2 for the largest $[2M+7H]^{7+}$ conformation. However, one should consider that its standard deviation is equal to 158 Å^2, whilst the experimental value is for a conformational family which was only observed once (Appendix 4).

---

* *corresponds to the range of dimer CCSs, displayed in Figure 4.24, at half peak height.*

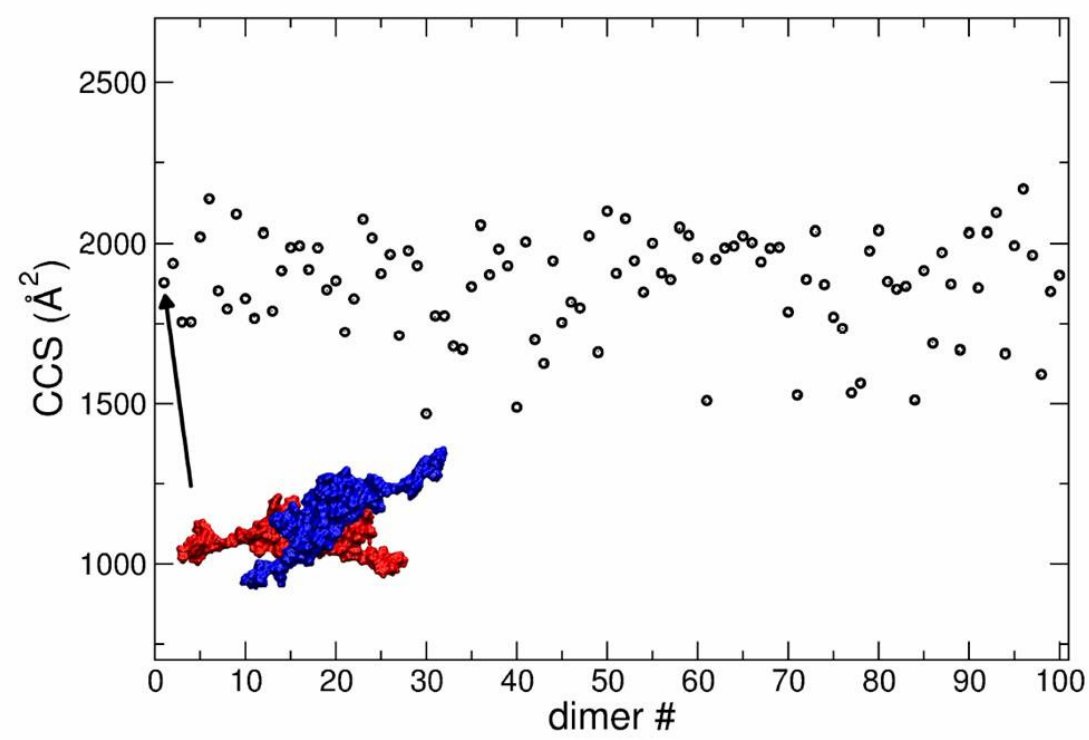


Figure 4.45 | CCS values of the dimers derived from the protein-protein docking between the structures with the highest R_g values. The dimer of the best docking is displayed with a surface representation with $[M+4H]^{4+}$ shown in blue and $[M+3H]^{3+}$ in red.

In order to reproduce the values of the most extended dimers, these very stretched monomeric structures have been manually joined in the four possible ways: N-N, C-C, N-C and C-N, where the C and N refer to the C-terminus and N-terminus respectively and the first letter of the pairs always refers to $[M+3H]^{3+}$. The structures are shown in Figure 4.46.

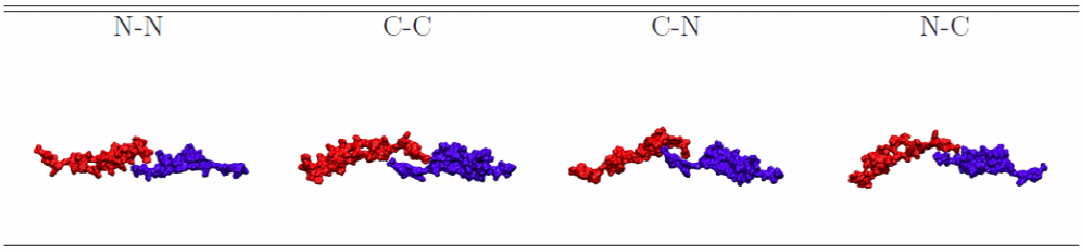


Figure 4.46 | The N-N, C-C, C-N and N-C docking structures in surface representation with $[M+4H]^{4+}$ shown in blue and $[M+3H]^{3+}$ in red.

Notably a geometry as extended as that observed experimentally for two conformational families of the $[2M+9H]^{9+}$ species might be explained by this end-on-end association on elongated monomers. CCS values calculated for the above structures were 2469.2 $Å^2$, 2406.7 $Å^2$, 2318.0 $Å^2$ and 2304.7 $Å^2$ respectively which compare favourably with the experimental $[2M+9H]^{9+}$ CCSs of 2305 ± 251 $Å^2$ and 2565 ± 72 $Å^2$.

4.4. Summary

MS allows observations of insulin oligomers up to dodecameric order before the onset of accelerated fibril growth, with dominant monomeric and dimeric species discerned. The additional separation power of IM-MS enables the identification of multiple conformations of insulin oligomers from dimeric to hexameric order and the observation of higher order aggromers composed of up to 24 monomers. Identities of multiple conformations have been confirmed by FT-ICR MS and are stable for milliseconds (IM-MS timescale) to seconds (FT-ICR MS timescale). This indicates that these species have a sufficiently high energy barrier to prevent inter-conversion during measurement. For all oligomeric species observed, the most significantly populated conformation is the most compact form. Modelling, for the first time trained by experimental measurements, has been used to characterise at atomistic detail aggregates bigger than a monomer. Modelling suggests that the most experimentally prevalent, conformationally compact, species are enriched in β-sheet secondary structure. Dimers of this form would fit the electron density previously found in single protofilaments of insulin.

4.5. References

1 Best, C. H. & Scott, D. A. The Preparation of Insulin. *J. Biol. Chem.* **57**, 709-723 (1923).

2 Brown, H., Sanger, F. & Kitai, R. The structure of pig and sheep insulins. *Biochem. J.* **60**, 556-565 (1955).

3 Adams, M. J. *et al.* Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224**, 491 (1969).

4 Jaimohan, S. M., Naresh, M.D., Mandal, A.B. Crystal structure of Bovine Insulin (2008) *not yet published*.

5 Brader, M. L. & Dunn, M. F. Insulin Hexamers - New Conformations and Applications. *Trends Biochem. Sci.* **16**, 341-345 (1991).

6 Fabris, D. & Fenselau, C. Characterization of allosteric insulin hexamers by electrospray ionization mass spectrometry. *Anal. Chem.* **71**, 384-387 (1999).

7 *NHS information website, Diabetes type 1*, http://www.nhs.uk/Conditions/Diabetes-type1/Pages/Introduction.aspx

8 *NHS information website, Diabetes type 2*, http://www.nhs.uk/Conditions/Diabetes-type2/Pages/Introduction.aspx

9 Brange, J. *et al.* Insulin structure and diabetes treatment. *Curr. Sci.* **72**, 470-476 (1997).

10 Howell, S. L., Tyhurst, M., Duvefelt, H., Andersson, A. & Hellerstrom, C. Role of Zinc and Calcium in Formation and Storage of Insulin in Pancreatic Beta-Cell. *Cell Tissue Res.* **188**, 107-118 (1978).

11 Brange, J., Andersen, L., Laursen, E. D., Meyn, G. & Rasmussen, E. Toward understanding insulin fibrillation. *J. Pharm. Sci.* **86**, 517-525 (1997).

12 Brange, J. & Volund, A. Insulin analogs with improved pharmacokinetic profiles. *Adv. Drug Del. Rev.* **35**, 307-335 (1999).

13 *Diabetes UK*, http://www.diabetes.org.uk/Guide-to-diabetes/Treatments/Insulin/

14 Nettleton, E. J. *et al.* Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophys. J.* **79**, 1053-1065 (2000).

15 Dische, F. E. *et al.* Inuslin as an Amyloid-Fibril Protein at Sites of Repeated Insulin Injections in a Diabetic Patient. *Diabetologia* **31**, 158-161 (1988).

16 Waugh, D. F. A Fibrous Modification of Insulin .1. The Heat Precipitate of Insulin. *J. Am. Chem. Soc.* **68**, 247-250 (1946).

17 Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317-1321 (2000).

18 Knowles, T. P. J. *et al.* An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **326**, 1533-1537 (2009).

19 Burke, M. J. & Rougvie, M. A. Cross-Beta Protein Structures .1. Insulin Fibrils. *Biochemistry* **11**, 2435 (1972).

20 Jahn, T. R. *et al.* The Common Architecture of Cross-β Amyloid. *J. Mol. Biol.* **395**, 717-727 (2010).

21 Tito, P., Nettleton, E. J. & Robinson, C. V. Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: a site-specific investigation by mass spectrometry. *J. Mol. Biol.* **303**, 267-278 (2000).

22 Nielsen, L. *et al.* Effect of environmental factors on the kinetics of insulin fibril formation: Elucidation of the molecular mechanism. *Biochemistry* **40**, 6036-6046 (2001).

23 Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M. & Robinson, C. V. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci.* **9**, 1960-1967 (2000).

24 McCullough, B. J. *et al.* Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal. Chem.* **80**, 6336-6344 (2008).

25 Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402-1407 (2002).

26 Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383-394 (2004).

27 Feig, M. *et al.* Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **25**, 265-284 (2004).

28 Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids*. (Clarendon Press, Oxford, UK, 1986).

29 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926-935 (1983).

30 Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an n.log(n) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).

31 Wang, G., Johnson, A. J. & Kaltashov, I. A. Evaluation of Electrospray Ionization Mass Spectrometry as a Tool for Characterization of Small Soluble Protein Aggregates. *Anal. Chem.* **84**, 1718-1724 (2011).

32 Du Vigneaud, V., Sifferd, R. H. & Sealock, R. R. The heat precipitation of insulin. *J. Biol. Chem.* **102**, 521-533 (1933).

33 Whittingham, J. L. *et al.* Insulin at pH 2: Structural Analysis of the Conditions Promoting Insulin Fibre Formation. *J. Mol. Biol.* **318**, 479-490 (2002).

34 Goto, Y., Calciano, L. J. & Fink, A. L. Acid-induced folding of proteins. *P. Natl. Acad. Sci. U. S. A.* **87**, 573-577 (1990).

35 Orci, L. *et al.* Conversion of proinsulin to insulin occurs coordinately with acidification of maturing secretory vesicles. *J. Cell Biol.* **103**, 2273-2281 (1986).

36 Waugh, D. F., Wilhelmson, D. F., Commerford, S. L. & Sackler, M. L. Studies of the Nucleation and Growth Reactions of Selected Types of Insulin Fibrils. *J. Am. Chem. Soc.* **75**, 2592-2600 (1953).

37 Bernstein, S. L. *et al.* Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature Chem.* **1**, 326-331 (2009).

38 Nettleton, E. J. Ph.D. thesis, Oxford University, (1998).

39 Jimenez, J. L. *et al.* The protofilament structure of insulin amyloid fibrils. *P. Natl. Acad. Sci. U. S. A.* **99**, 9196-9201 (2002).

40 Bleiholder, C., Dupuis, N. F., Wyttenbach, T. & Bowers, M. T. Ion mobility mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. *Nature Chem.* **3**, 172-177 (2011).

41 Kloniecki, M. *et al.* Ion Mobility Separation Coupled with MS Detects Two Structural States of Alzheimer's Disease Aβ1-40 Peptide Oligomers. *J. Mol. Biol.* **407**, 110-124 (2011).

42 Smith, D. P., Radford, S. E. & Ashcroft, A. E. Elongated oligomers in β_2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *P. Natl. Acad. Sci. U. S. A.* **107**, 6794-6798 (2010).

43 Clemmer, D. E. & Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass Spectrom.* **32**, 577-592 (1997).

44 Jarrold, M. F. Peptides and proteins in the vapor phase. *Annu. Rev. Phys. Chem.* **51**, 179-207 (2000).

45 Fernandez de la Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93-104 (2000).

46 Ahmad, A., Uversky, V. N., Hong, D. & Fink, A. L. Early events in the fibrillation of monomeric insulin. *J. Biol. Chem.* **280**, 42669-42675 (2005).

47 Ahmad, A., Millett, I. S., Doniach, S., Uversky, V. N. & Fink, A. L. Partially folded intermediates in insulin fibrillation. *Biochemistry* **42**, 11404-11416 (2003).

48 http://biophysics.cs.vt.edu/H++/

49 Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**, 16082-16086 (1996).

50 Amber 10 (University of California, San Francisco, 2008).

51 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).

52 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802 (2005).

53 Shao, J. Y., Tanner, S. W., Thompson, N. & Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **3**, 2312-2334 (2007).

54 http://hex.loria.fr/

55 Breuker, K. & McLafferty, F. W. Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^{2} S. *P. Natl. Acad. Sci. U. S. A.* **105**, 18145-18152 (2008).

56 Kabsch, W. & Sander, C. Dictonary of Protein Secondary Structure - Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **22**, 2577-2637 (1983).

57 Kollman, P. A. *et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**, 889-897 (2000).

58 Gilson, M. K. & Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21-42 (2007).

59 Barran, P. E. *et al.* Is it biologically relevant to measure the structures of small peptides in the gas-phase? *Int. J. Mass spectrom.* **240**, 273-284 (2005).

5

Aβ Peptides and Inhibitors

Aβ peptides Aβ(1-40) and Aβ(1-42) are implicated in the pathogenic mechanisms of Alzheimer's Disease. In this chapter, MS (in both positive and negative ionisation modes) and IM-MS are utilised to analyse Aβ peptide oligomer populations and the complexes they form with the retero-inverso OR2 peptide, which has been reported to be an inhibitor of Aβ oligomerisation and fibril formation. IM-MS results elucidate a compaction of Aβ peptides in complex with retero-inverso OR2, which could be the mode of action of this inhibitor.

5.1. Introduction

Amyloid β (Aβ) peptides are formed by the proteolytic cleavage of the transmembrane amyloid β precursor protein (APP) [1], a cell traffic regulating glycoprotein [2] which is expressed at high levels in the mammalian brain. The action of a β secretase outside the membrane, followed by a γ secretase transmembrane cleavage, produces an Aβ peptide of between 36 and 43 amino acids in length (Figure 5.1). The peptides Aβ(1-40) and Aβ(1-42) are the most frequent products of this cleavage [3] and are amyloidogenic in nature [4]. An alternative, non-amyloidogenic, pathway occurs when APP is cleaved by an α secretase [5], followed by the same γ secretase transmembrane cleavage. The peptides produced by this succession of secretase cleavages do not form oligomers or amyloid fibrils [6].

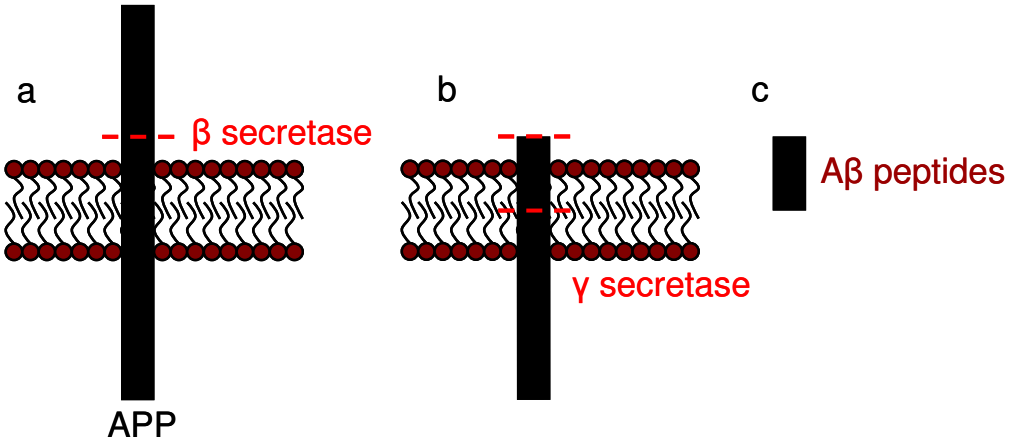


Figure 5.1 | Schematic of the production of Aβ peptides from APP by the action of secretase enzymes. a. Cleavage of APP by β secretase followed by b. γ secretase transmembrane cleavage producing c. Aβ peptides

The localisation of Aβ(1-40) and Aβ(1-42) production is different; in neurons, Aβ(1-40) creation occurs in the endoplasmic reticulum, whilst Aβ(1-42) is formed in the trans-Golgi network [7]. Although 10 times less numerous than Aβ(1-40) *in vivo*, Aβ(1-42) is strongly linked to the aetiology of, and is more abundant in, the amyloid plaques which are diagnostic of Alzheimer's disease [8]. Elevated Aβ(1-42) levels in plasma are also correlated with the development of Alzheimer's disease [9]. Aβ(1-42)

is comparatively more cytotoxic to neurons than Aβ(1-40) [10-13] and *in vitro* has a higher aggregation propensity [14]. Aβ(1-40) and Aβ(1-42) fibrils both possess the cross-β structure shared by all amyloid fibrils [15] and have β-turn-β motifs which are of similar conformation with parallel stacking and an in-register configuration [16]. Figure 5.2 summarises the current models for the β-turn-β motifs present in Aβ(1-40) and Aβ(1-42) fibrils.

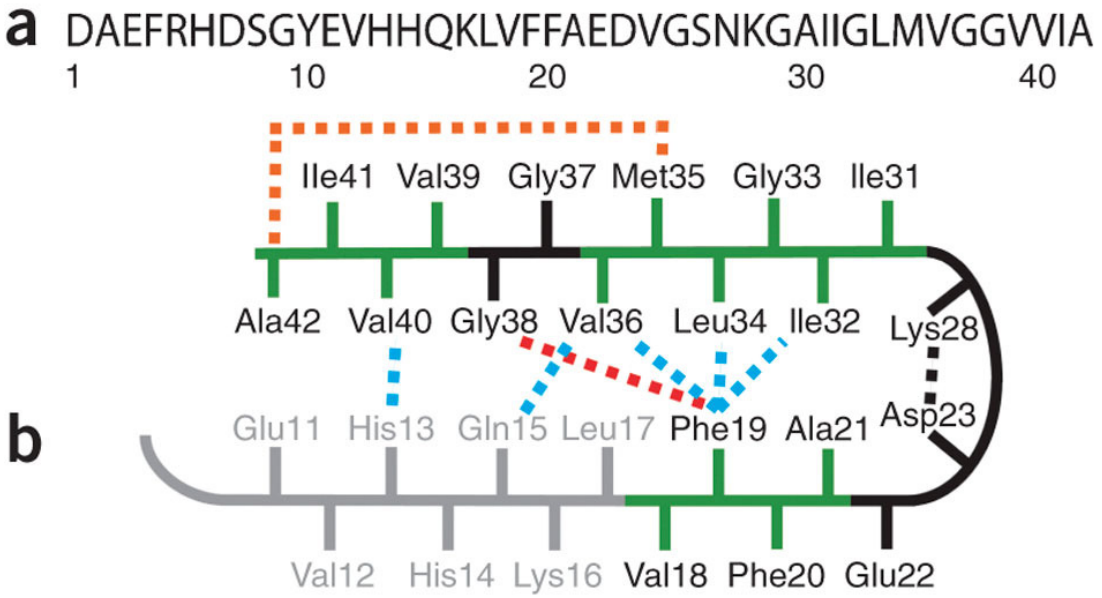


Figure 5.2 | Schematic of the central residues of Aβ peptides (a) amino acid sequence of Aβ(1-42) (b) Structural constraints in Aβ(1-40) and Aβ(1-42) fibrils. SSNMR measurements of Aβ(1-40) fibrils suggest side chain packing between Phe19 and Ile32, Leu34 and Val36, Gln15 and Val36 and His13 and Val40 (blue dashed lines). Molecular contacts have been reported within the monomer unit of Aβ(1-42) fibrils between Phe19 and Gly38 (red dashed line) [17] and between Met35 and Ala42 (orange dashed line) [18]. In both Aβ(1-40) and Aβ(1-42), the turn conformation is stabilised by hydrophobic interactions (green residues) and by a salt bridge between Asp23 and Lys28 (black dashed line). Figure and caption adapted from Ahmed et al. [16].

SSNMR studies combined with computational approaches have allowed a structural model of lyophilised mature Aβ(1-40) fibrils, formed at pH 7.4, to be proposed. The core region of the fibrils is formed by two β-strands from each Aβ(1-40) monomer, a β-turn-β fold, comprising residues 12 to 24 and 30 to 40 [19,20]. The Aβ(1-40) monomers are stacked predominantly in an in-register and parallel alignment [21-23]. Aβ(1-40) N-termini are flexible and solvent exposed [24-26]; specifically residues 1 to 10 are unstructured [19,20]. TEM of Aβ(1-40) fibrils reveals morphological differences dependant on agitation or quiescence, with SSNMR reporting that different molecular structures and intermolecular contacts within the fibril cause polymorphism [27]. Fibrillar Aβ(1-42) is reported to possess an unstructured N-terminus up to residue 10 [16,17,28] and a lack of structure in the last two or three amino acids of the C-terminus [29]. The central Aβ(1-42) residues compose two parallel and in-register β strands [23,29,30], with residues 18 to 42 composing the β-turn-β configuration [17]. The remainder of the C-terminal residues are protected from solvent exchange [17,28]. As mentioned in Chapter One, there is confusion in the literature as to the nomenclature and identity of prefibrillar intermediates. Observations of Aβ protofibrils by AFM and TEM are varied, including 2 to 5 nm spheres, connection of these spheres into beaded chains and closed chain rings [31-34]. Aβ protofibrils exhibit binding with Congo Red and Thioflavin T dyes, diagnostic of the presence of β-sheet structure [34]. Proline scanning mutagenesis has been used to elucidate Aβ protofibril structure discovering similar flexible N and C-termini to those of mature fibrils, but a lower level of structure in residues 22 to 29 [35].

Aβ(1-40) and Aβ(1-42) exist *in vitro* as monomers prior to their aggregation into soluble oligomers and amyloid fibrils [36]. The monomeric conformations of Aβ(1-40) and Aβ(1-42) are slightly different, with Aβ(1-42) possessing greater β-sheet and β-turn structure [37]. Soluble oligomer protofibril precursors, which are in rapid equilibria with the monomers, have been observed for both Aβ(1-40) and Aβ(1-42), *via* photo-induced cross linking [38]. Aβ(1-40) and Aβ(1-42) are postulated to participate in differing assembly pathways; Bitan *et al.* report that Aβ(1-40) forms an equilibrium of monomers, dimers, trimers and tetramers [37], whereas Aβ(1-42) has been proposed to rapidly form structurally disorganised pentamers and hexamers

which act as 'paranuclei', associating into larger oligomer and protofibrils [37]. This is supported by a recent paper by Ahmed *et al.* [16] which reports Aβ(1-42) oligomers of mostly pentameric and hexameric order but a small population of decamers and dodecamers, which they postulate to have pentameric and hexameric building blocks. These Aβ(1-42) oligomers, which contained no β-sheet structure, were more toxic to murine neurons than Aβ(1-42) protofibrils or mature fibrils [16]. This is of particular interest as small oligomers are found in the brains of Alzheimer's patients [39] and in *in vitro* studies [40,41]. Thus there is accumulating evidence that Aβ oligomers, particularly the Aβ(1-42) decamer and dodecamer [42-44], are the principle pathological species in the progression of Alzheimer's disease [45,46], for example; Aβ oligomers are reported to cause neuron dysfunction [47-49] and murine memory loss [42]. The additional two C-terminal residues present in Aβ(1-42) result in a marked increase in aggregation propensity and cytotoxicity. Ahmed *et al.* [16] suggest that this is due to the additional hydrophobic residues in Aβ(1-42) increasing the stability of neurotoxic oligomers.

5.1.1. Aβ MS, IM-MS and Modelling

Grasso has recently written a comprehensive review of the use of MS to study Aβ peptides [36], therefore this brief review focuses on studies which have utilised IM-MS and molecular modelling.

IM-MS, in conjunction with molecular modelling, has been used to investigate Aβ(1-42) monomer structure [50]. Simulated Aβ(1-42) monomers in solution were found to be conformationally dynamic and could be classified into three conformational families. Secondary structure calculated for these monomers *in silico* consisted of mostly turns and loops with some helical elements in the C-terminus. A combination of calculated CCSs of solvent free structures and theoretical CCSs of the monomer families after dehydration, agreed well with the experimental ATD of the most dominant mass spectral peak $[Aβ(1-42)-3H]^{3-}$. The advantage conferred by using IM-MS to elucidate the early stages of the aggregation process is clear as

demonstrated by Bernstein *et al.* [51] in the comparison of oligomeric species of Aβ(1-42) and the non-amyloidogenic variant Aβ(1-42) F19P. The mass spectral peak corresponding to the $[2M-5H]^{5-}$ species yielded an ATD which suggested differences between the WT and mutant peptides (Figure 5.3a and b). After injection energy studies the ATD peaks were assigned as dimeric and tetrameric species for Aβ(1-42) F19P, and dimer, tetramer, hexamer and dodecamer species for Aβ(1-42).

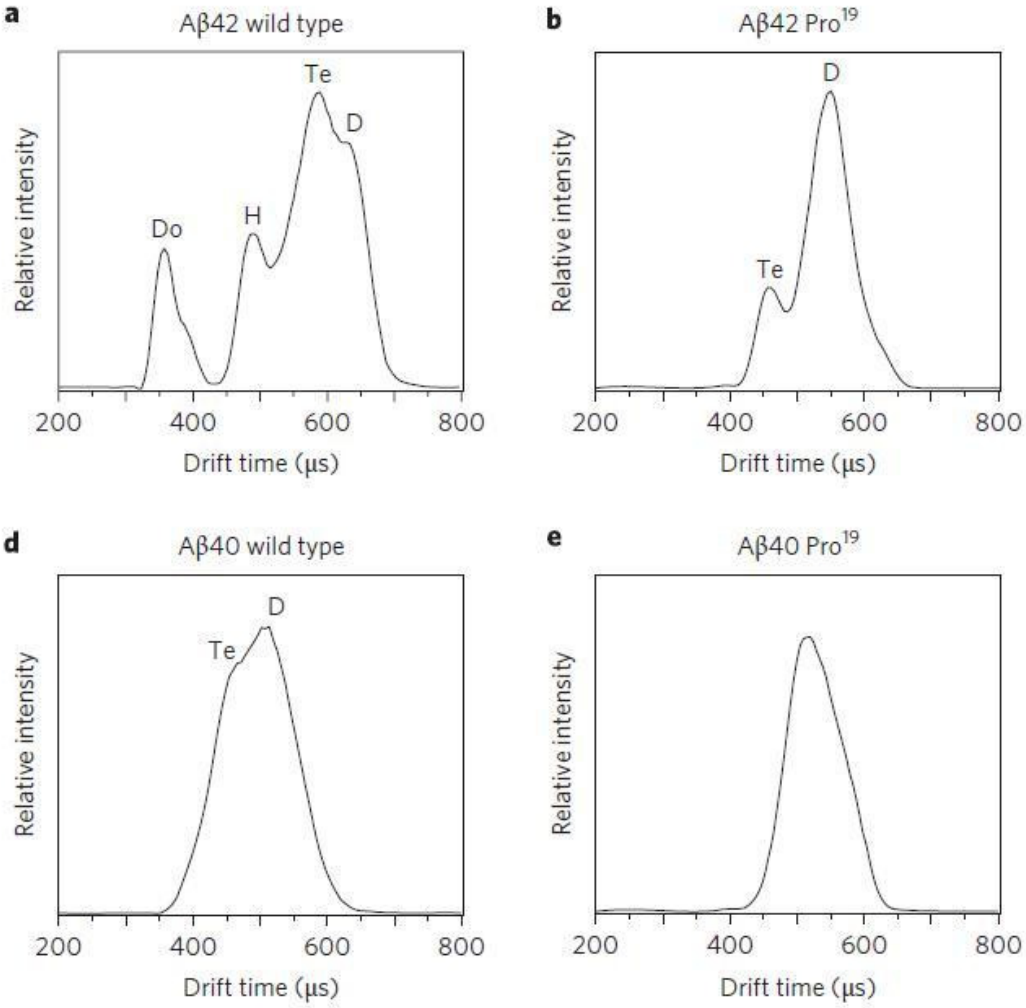


Figure 5.3 | ATDs of $[2M-5H]^{5-}$ for Aβ(1-42) (a,b) and Aβ(1-40) (d,e), showing the wild type (a,d) and mutation F19P (b,e). D denotes a peak assigned as a dimer, T a tetramer, H a hexamer and Do dodecamer. Figure and caption from Bernstein et al. [52].

The ATD of WT Aβ(1-40) more closely resembled that of Aβ(1-42) F19P than that of WT Aβ(1-42). This led the authors to propose that Aβ(1-42) possessed the ability to form a more open tetramer than other Aβ peptides and Aβ(1-42) alloforms [52]. It is hypothesised that these open tetramers could associate with dimers to form hexamers, which stack together to form dodecamers; a species capable of beginning the fibril assembly process, a finding supported by cross-linking studies [37,38,53]. Bernstein *et al.* [52] also postulated that this dodecameric species was the soluble species implicated by Lesné *et al.* [42] in murine memory impairment. When Aβ(1-40) is present with Aβ(1-42) in solution, Aβ(1-42) oligomerisation is inhibited, as the largest oligomeric species formed are mixed Aβ(1-42):Aβ(1-40) tetramers which appear not to be growth competent, possibly due to their compact nature, thus preventing the formation of the putative toxic dodecamer [54].

MS and IM-MS Experiments performed on Aβ(1-40) have yielded conflicting results. Negative ionisation mode MS and IM-MS revealed oligomers up to tetrameric order [52], however a positive ionisation mode MS and IM-MS study undertaken by Kloniecki *et al.* [55], using a different instrument, yielded oligomers composed of up to 16 monomers [55]. Differing Aβ(1-40) preparations and increased instrument sensitivity in the later study are suggested as the origin for this difference [55]. However the presence of higher order oligomers, including a dodecameric Aβ(1-40) species, has serious implications for the self assembly pathways postulated by Bernstein *et al.* [52] (described above). MS and IM-MS experiments conducted in this chapter attempt to resolve this conflict.

5.1.2. Aβ Aggregation Inhibitors

Aβ peptide residues 16 to 22, KLVFFAE, are believed to play a vital role in the aggregation process [56,57]. Extensive research has been carried out to design inhibitors with recognition sequences that can target this region. One class of inhibitors are known as β breakers. These are short peptides which bind to the target region and discourage the formation of β-sheet secondary structure [58]. In the Aβ context, β

breakers contain a proline residue, which, when it binds to the target sequence, creates a kink and breaks the β-sheet hydrogen bonding pattern (Figure 5.4) [59,60]. β breakers have been reported to inhibit the creation of mature fibrils, although their effect on the formation of oligomers, cytotoxic or otherwise, is unknown [59,60].

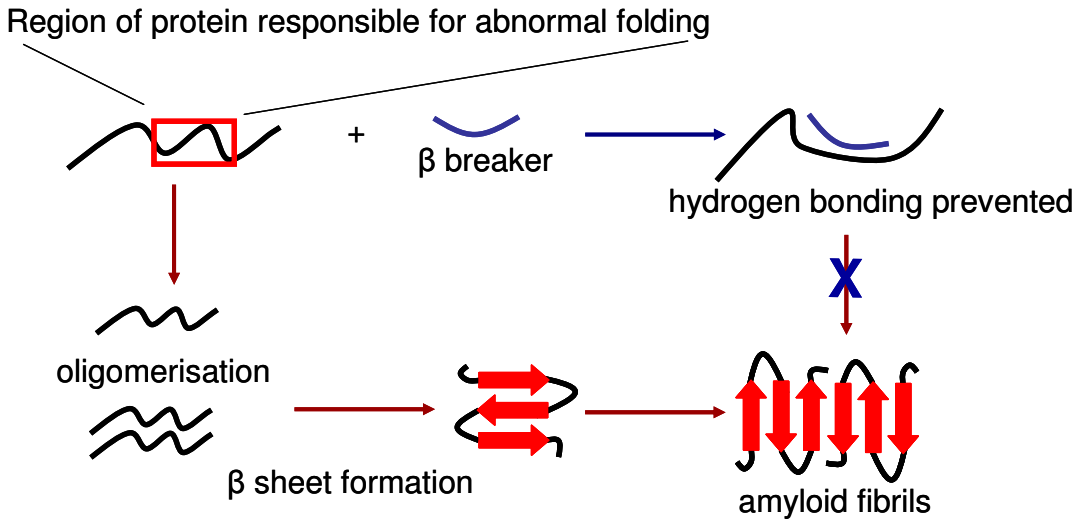


Figure 5.4 | Schematic of β breaker peptides in action.

There are three classes of Aβ aggregation inhibitors, as defined by Necula *et al.* [61], these are shown in Table 5.1.

Aβ Aggregation Inhibitor	Action
Class I	Inhibit oligomerisation
Class II	Inhibit the formation of both oligomers and fibrils
Class III	Inhibit fibril formation but not oligomerisation

Table 5.1 | Table of Aβ aggregation inhibitors.

5.1.2.1. Retro Inverso OR2

Retro-inverso OR2 is a peptide based on OR2, a known Aβ oligomer and fibril formation inhibitor [62]. OR2 has the amino acid sequence H_2N-Arg-Gly-Lys-Leu-Val-Phe-Phe-Gly-Arg-NH_2, which mimics the Aβ residue sequence 16-20 to encourage

binding. OR2 is a modified version of OR1, possessing an amidated C-terminus. OR1 has been reported to prevent the formation of mature Aβ fibrils without influencing oligomer formation [62]. OR2 contains additional amino acids to those which comprise the binding sequence Lys-Leu-Val-Phe-Phe to increase functionality. The presence of cationic arginine residues at both termini increases solubility and discourages self-aggregation, whilst glycine spacers [62] aid interactions between Aβ and the binding sequence region of OR2. However, OR2 is vulnerable to proteolytic cleavage which limits its potential as a therapeutic. To design a peptide with similar activity (and therefore the need to preserve the three dimensional shape of OR2) which was resistant to proteolytic attack, the L-amino acids of OR2 were substituted with D-enantiomers and all peptide bonds reversed [63,64]. This resulted in the retro-inverso version of OR2 (RI-OR2) with sequence: H_2N-r←G←k←l←v←f←f←G←r-Ac [65]. RI-OR2 was found to be resistant to proteolytic degradation by a variety of enzymes, all of which caused degradation of OR2. RI-OR2 also possessed a much higher stability in blood plasma and human brain extract than OR2, further validating the retro-inverso approach [65].

Analysis of thioflavin T studies indicate that an equimolar quantity of RI-OR2 blocks the process of fibril formation by Aβ(1-40) and Aβ(1-42) more effectively than an equimolar quantity of OR2. However, at substoichiometric concentrations of inhibitor OR2 appeared to perform better. Both OR2 and RI-OR2 inhibit Aβ(1-42) aggregation more effectively than that of Aβ(1-40) [65], not a disadvantage given the higher amyloidogenicity of Aβ(1-42) than Aβ(1-40). Oligomer formation by both forms of Aβ was monitored using a sandwich immunoassay [62] and reported to be prevented by both OR2 and RI-OR2 at inhibitor:Aβ ratios of 2:1, 1:1 and 1:2 with increased efficacy against Aβ(1-42) oligomer formation [65]. Additional experiments utilised SDS page and immunodetection to determine that Aβ(1-42) oligomers of nonameric to dodecameric order and dimeric to tetrameric order, which were present in an aggregating control sample, were not present in a 1:1 stoichiometric mix of RI-OR2:Aβ(1-42). OR2 and RI-OR2 were also found to protect against the cytotoxicity of aggregated Aβ(1-42) [65]. Surface plasmon resonance experiments showed that RI-OR2 binds to Aβ(1-42) monomers and fibrils quickly, with a binding

affinity of 9 μM and 12 μM respectively [65]. Whereas OR1 is a class III inhibitor, inhibiting fibrillisation but not oligomerisation, both OR2 and RI-OR2 can be classified as class II inhibitors [61] as they prevent oligomerisation, fibril formation and extension. This class of inhibitors are postulated to stabilise a non-aggregating conformation of Aβ, which RI-OR2 achieves by binding to the KLVFF motif. Taylor *et al.* [65] suggest that Aβ(1-42) monomer binding is the means by which oligomer formation and aggregation is prevented. Aβ(1-40) and Aβ(1-42) have slightly different monomeric conformations (as described previously) [37] and thus the interaction of RI-OR2 to monomeric Aβ(1-42) is either stronger or it acts more effectively on the Aβ(1-42) assembly process.

5.2. Methodology

5.2.1. Sample Preparation

Inhibitor RI-OR2 was obtained as a lyophilised powder from Mark Taylor at Lancaster University [*]. Recombinant Aβ peptides were purchased from rPeptide (Bogart, USA) and arrived pre-treated with hexafluoro-2-propanol. Aβ peptides were deseeded by dissolving in trichloroacetic acid and thioanisole, incubating for an hour, and the liquid removed in a nitrogen stream. The residue is then dissolved in hexafluoro-2-propanol and evaporated twice before being split into 0.1 mg aliquots [†].

To create an aggregating solution the peptide was dissolved at a concentration of 100 μM in H_2O adjusted to pH 2 with hydrochloric acid. These samples were immediately flash frozen in liquid nitrogen and stored in a -28 °C freezer.

[*] Division of Biomedical and Life Sciences, School of Health and Medicine, Lancaster University, Lancaster LA1 4YQ, UK.
[†] Deseeding procedure carried out by Mark Taylor at Lancaster University.

In initial negative ionisation mode experiments, samples were ziptipped to remove Cl^- adducts. Ziptipping utilised Millipore ZipTip® C_{18} pipette tips, the procedure is described below:

Ziptipping Procedure:

1. Wetting solution of 100 % ACN: repeat x 2.
2. Equilibrium solution of 0.1 % TFA in milli-Q water: repeat x 2.
3. Peptide solution: repeat x 10.
4. Wash solution of 5 % MeOH in 0.1 TFA in milli-Q water: repeat x 2.
5. Elution solution of 1 % HCOOH in 50 % aqueous MeOH: repeat x 4.

Solution conditions following ziptipping were therefore 49.5:49.5:1 MeOH:H_2O:HCOOH.

5.2.2. Mass Spectrometry

High resolution mass spectrometry was performed on a QTOF 2 mass spectrometer (Waters, Manchester, UK) in both positive and negative ionisation modes. High resolution mass spectra provided ^{13}C isotopic distributions which confirmed the identification of peaks where several assignations might have been made due to the presence of oligomers with a coincident *m/z*.

5.2.3. Ion Mobility Mass Spectrometry

IM-MS measurements were performed on an in-house modified QTOF 1 (Waters, Manchester, UK) the 'MoQTOF' [66] (Chapter 2, Section 2.3.1). Samples were ionised using an n-ESI source in positive mode with an applied capillary voltage of ~1.8 kV and a cone voltage of 40 V. Ions were accelerated through a potential difference of 36 V and injected into the drift cell. The drift cell was filled with He gas at a pressure of between 3.0 and 3.5 Torr and a temperature between 300 and

305 K. The temperature and pressure of helium was recorded at the start and end of each mobility scan and the average taken. The voltage applied across the cell was varied from 60 V to 15 V and ATDs were recorded at eight voltages.

5.2.4. Collision Induced Dissociation Experiments

Collision induced dissociation (CID) experiments in an argon filled collision cell were performed using a QTOF 2 (Waters, Manchester, UK). Peaks were selected by their *m/z* value and accelerated into the collision cell at a variety of voltages. The kinetic energy of the ions entering the collision cell was increased by raising the collision voltage parameter until the signal from the parent ion had been completely lost and only product ions remained.

5.3. Results and Discussions

5.3.1. Mass Spectrometry of Aβ(1-40), Aβ(1-42) and RI-OR2

For the purpose of clarity in mass spectrometry notation, Aβ(1-40) and Aβ(1-42) are referred to as Aβ40 and Aβ42 respectively. Negative ionisation MS experiments were attempted with unziptipped samples but were unsuccessful due to suppression of sample signal by chloride adduct peaks. Therefore all data acquired in negative ionisation mode utilised ziptipped samples (procedure detailed in Section 5.2.1).

5.3.1.1. Negative Ionisation Mode

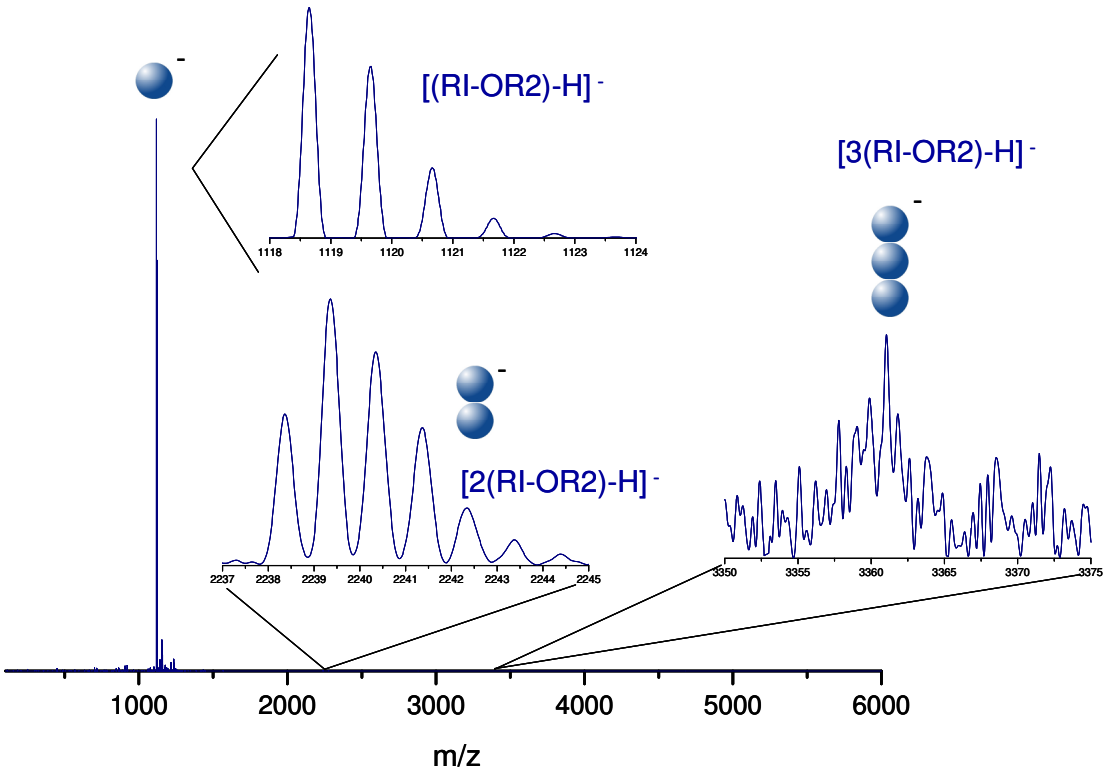


Figure 5.5 | Mass spectra of ~50 µM ziptipped RI-OR2 in pH 2 49.5:49.5:1 MeOH:H_2O:HCOOH.

As the amino acid sequence of the RI-OR2 peptide contains no aspartic or glutamic acid residues, and thus possesses a low gas phase acidity, it is unsurprising that the species observed are not more than singularly charged. The inclusion of arginine residues at both termini of RI-OR2 to discourage self-aggregation [62] has been successful, as from Figure 5.5 a dominant $[(RI\text{-}OR2)\text{-}H]^-$ species can be discerned, with an extremely small amount of dimer and trimer present. For a 100 % intensity of the $[(RI\text{-}OR2)\text{-}H]^-$ species, dimeric $[2(RI\text{-}OR2)\text{-}H]^-$ has an intensity of 0.04 % and $[3(RI\text{-}OR2)\text{-}H]^-$ an intensity of 0.002 %.

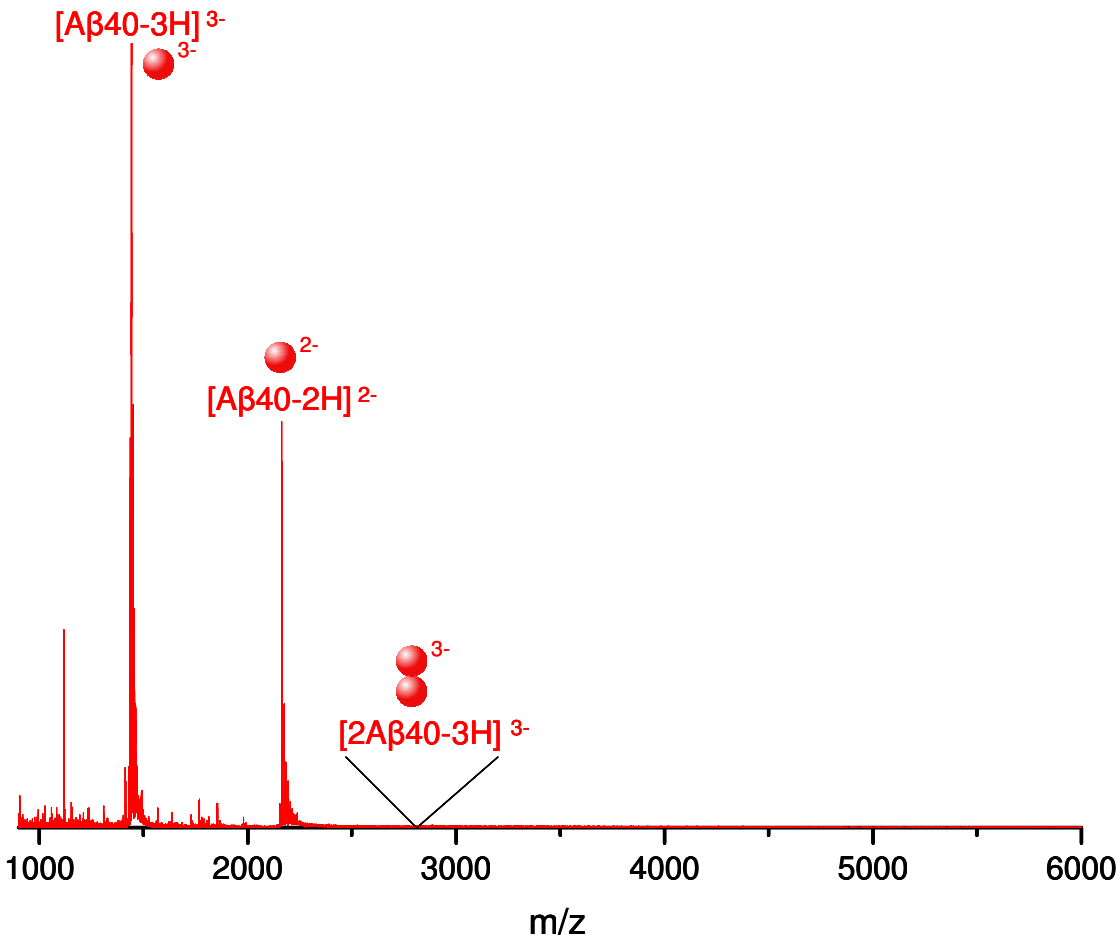


Figure 5.6 | Mass spectrum of ~50 μM ziptipped Aβ(1-40) in pH 2 49.5:49.5:1 MeOH:H_2O:HCOOH.

The peak corresponding to $[Aβ40\text{-}3H]^{3-}$ dominates the mass spectrum displayed in Figure 5.6, in addition there is a significant contribution from the $[Aβ40\text{-}2H]^{2-}$ species. The most populated oligomeric species observed is $[2Aβ40\text{-}3H]^{3-}$; other less abundant species which are present but not visible in Figure 5.6 are $[Aβ40\text{-}4H]^{4-}$, $[2Aβ40\text{-}7H]^{7-}$, $[3Aβ40\text{-}7H]^{7-}$, $[3Aβ40\text{-}8H]^{8-}$ and $[4Aβ40\text{-}6H]^{6-}$. Negative ionisation mode MS in the literature [52] also reports a dominant $[Aβ40\text{-}3H]^{3-}$ species, followed by mass spectral peaks corresponding to $[Aβ40\text{-}4H]^{4-}$, $[2Aβ40\text{-}5H]^{5-}$ and $[Aβ40\text{-}2H]^{2-}$, however no oligomers of greater than dimeric order were discerned *via* MS with an unique *m/z*. The difference in Aβ(1-40) charge state distribution can be

accounted for by the different pH (7.4) and solution conditions (25 mM ammonium acetate) employed by Bernstein *et al.* [52], which favour higher negative charge states of Aβ(1-40).

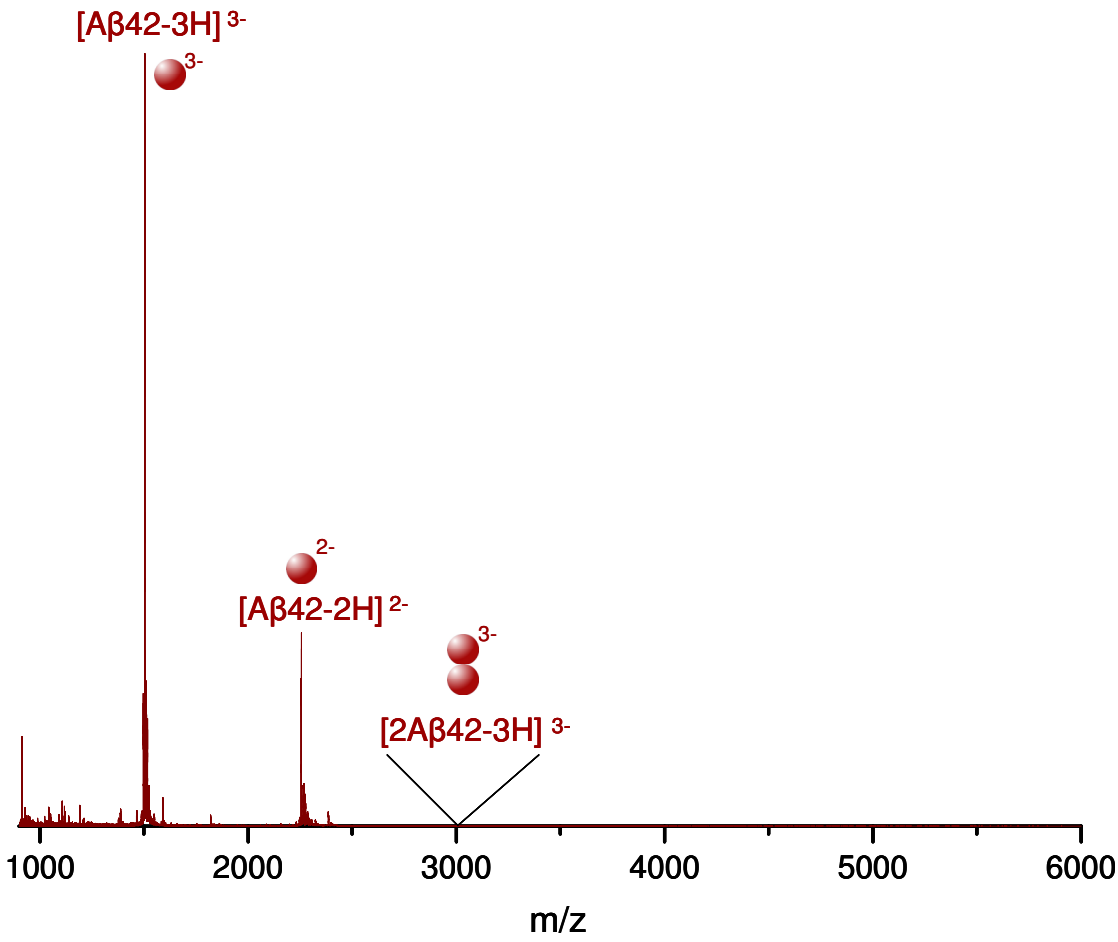


Figure 5.7 | Mass spectrum of ~50 μM ziptipped Aβ(1-42) in pH 2 49.5:49.5:1 MeOH:H_2O:HCOOH.

Despite the similarity of the dominant mass spectral peaks displayed in Figure 5.6 and Figure 5.7, enlargements of the mass spectrum of Aβ(1-42) reveal a paucity of oligomers compared to Aβ(1-40), under identical solution and experimental conditions. Figure 5.7 displays a dominant peak corresponding to $[Aβ42\text{-}3H]^{3-}$, a less intense $[Aβ42\text{-}2H]^{2-}$ signal and small peak corresponding to the presence of $[2Aβ42\text{-}3H]^{3-}$. This is a similar charge state distribution to that of Aβ(1-40) (Figure

5.6) however there is only one dimeric species, and no trimers or tetramers detected. Bernstein *et al.* [51] report a dominant $[A\beta42\text{-}3H]^{3-}$ species, followed by, in decreasing signal intensity: $[A\beta42\text{-}4H]^{4-}$, $[A\beta42\text{-}2H]^{2-}$, $[2A\beta42\text{-}5H]^{5-}$ and $[A\beta42\text{-}5H]^{5-}$. This disparity in Aβ(1-42) charge state distribution can be accounted for by the differing pHs of the two experiments, with those conducted by Bernstein *et al.* at a more basic pH (near 8) which, in negative ionisation mode, favours a shift towards higher Aβ(1-42) charge states.

5.3.1.2. Interactions between Aβ Peptides and RI-OR2

Aβ peptides were mixed with RI-OR2 to observe the interactions between the peptides and the inhibitor.

As can be observed in Figure 5.8, the $[(RI\text{-}OR2)\text{-}H]^-$ species dominates the mass spectrum and, as a result, Aβ(1-40) charge states are only visible on enlargement of their *m/z* regions. The distribution of Aβ(1-40) charge states (not visible in Figure 5.8) matches that observed when Aβ(1-40) is analysed in isolation (Figure 5.6). However, additional peaks corresponding to the presence of Aβ(1-40) and RI-OR2 complexes are present. Complexes $[A\beta40+(RI\text{-}OR2)\text{-}3H]^{3-}$, $[A\beta40+(RI\text{-}OR2)\text{-}2H]^{2-}$, $[A\beta40+2(RI\text{-}OR2)\text{-}2H]^{2-}$ and $[A\beta40+2(RI\text{-}OR2)\text{-}3H]^{3-}$ are observed and are displayed in the insets of Figure 5.8.

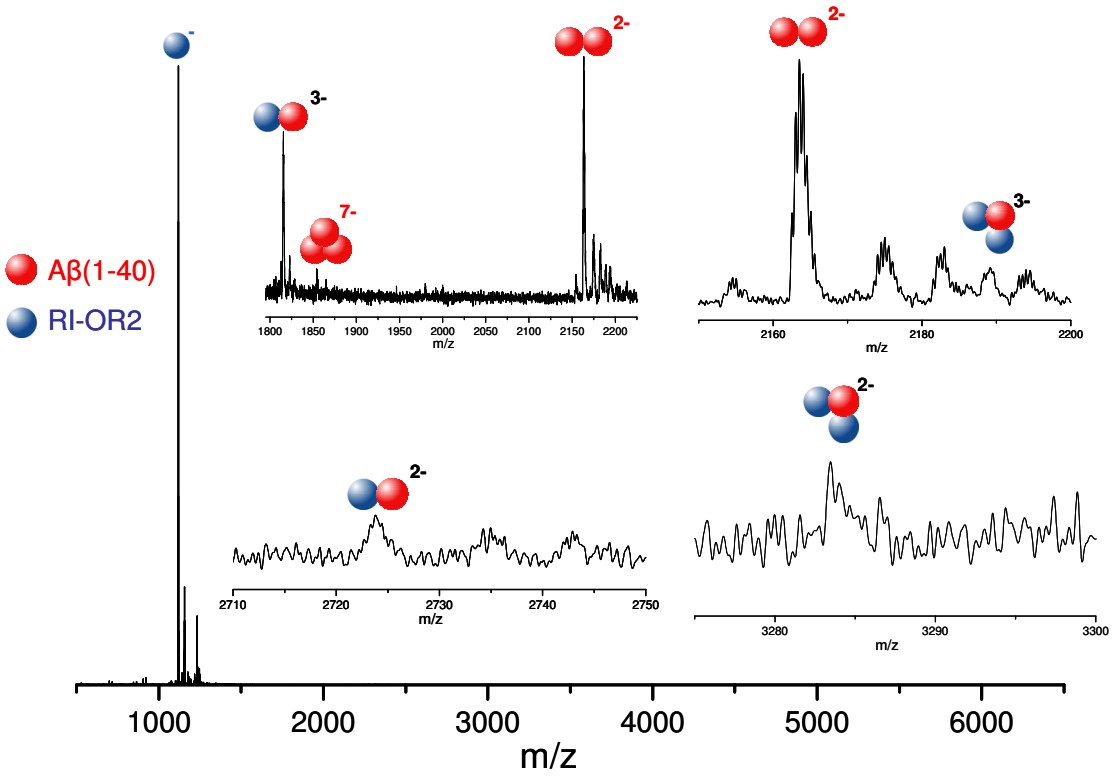


Figure 5.8 | Mass spectra of ~50 μM ziptipped Aβ(1-40) and RI-OR2 in an ~1:1 mixture. Solvent conditions are 49.5:49.5:1 MeOH:H_2O:HCOOH at pH 2.

$[(RI\text{-}OR2)\text{-}H]^-$ is again the dominant species in the mass spectrum displayed in Figure 5.9. The Aβ(1-42) charge states are only visible on enlargement of their *m/z* regions, but their distribution matches that observed when Aβ(1-42) is analysed in isolation (Figure 5.7). Complexes of Aβ(1-42) and RI-OR2 are present: $[Aβ42+(RI\text{-}OR2)\text{-}3H]^{3-}$, $[Aβ42+(RI\text{-}OR2)\text{-}2H]^{2-}$, $[Aβ42+2(RI\text{-}OR2)\text{-}2H]^{2-}$ and $[Aβ42+2(RI\text{-}OR2)\text{-}3H]^{3-}$. These are displayed in the inset of Figure 5.9. The relative abundances of the Aβ and RI-OR2 complexes is similar for both Aβ(1-40) and Aβ(1-42); $[Aβ+(RI\text{-}OR2)\text{-}3H]^{3-}$ is the most populated complex present, followed by $[Aβ+2(RI\text{-}OR2)\text{-}3H]^{3-}$, $[Aβ+(RI\text{-}OR2)\text{-}2H]^{2-}$ and the least populated complex $[Aβ+2(RI\text{-}OR2)\text{-}2H]^{2-}$.

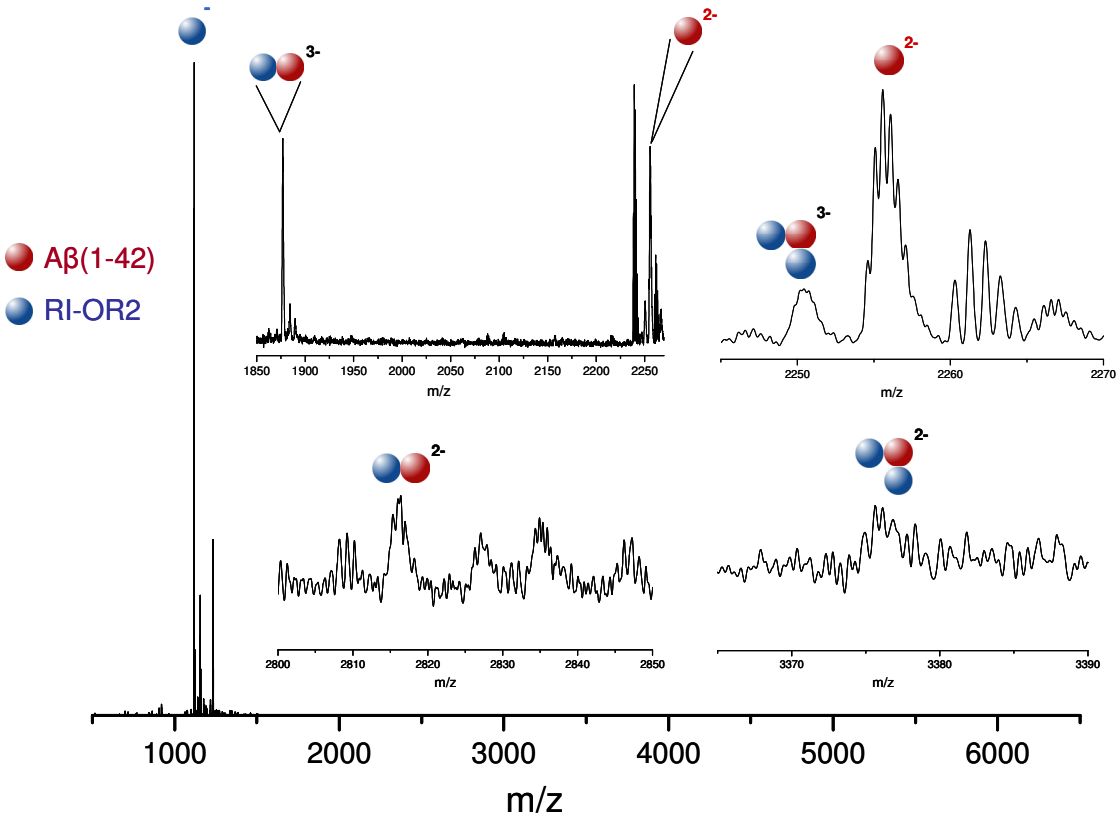


Figure 5.9 | Mass spectra of ~50 μM ziptipped Aβ(1-42) and RI-OR2 in an ~1:1 mixture. Solvent conditions are 49.5:49.5:1 MeOH:H_2O:HCOOH at pH 2.

Due to the low intensities of some of the complex peaks identified in the 1:1 Aβ:RI-OR2 solutions, the mass spectra of which are displayed in Figure 5.8 and Figure 5.9, CID experiments were performed to verify their presence. The results of these experiments and the corresponding dissociation pathways are shown in Figure 5.10.

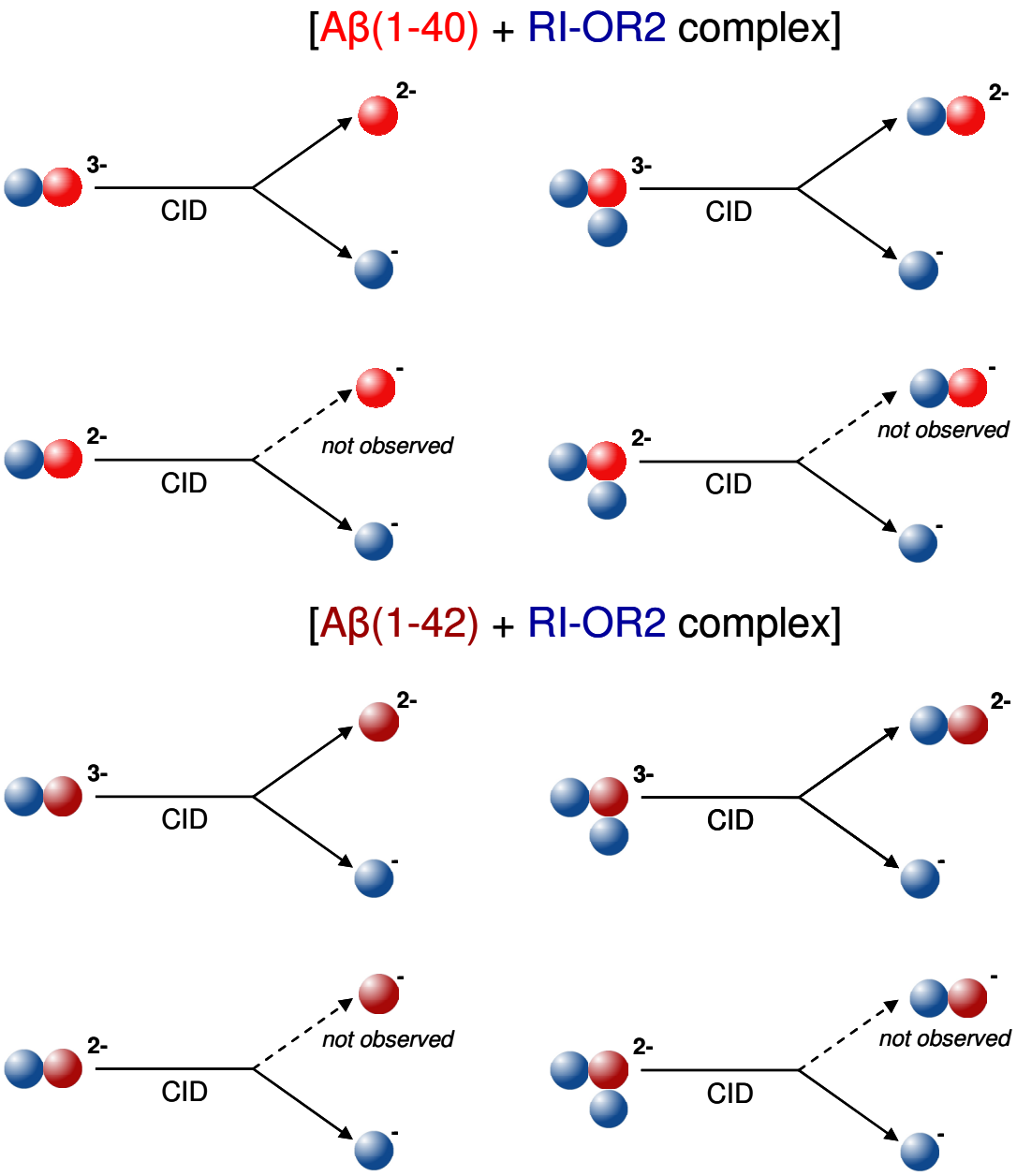


Figure 5.10 | Schematic of the dissociation pathways of RI-OR2 and Aβ complexes. Dashed lines represent product species which were not experimentally observed.

CID verified that all complexes had been correctly identified. Figure 5.10 shows that the dissociation of complexes of Aβ(1-42) with RI-OR2 follow similar pathways to that of complexes composed of Aβ(1-40) and RI-OR2. The dissociation pathways of all complexes between Aβ and RI-OR2 produce a product $[(RI\text{-}OR2)\text{-}H]^-$ ion. The

prevalence of $[(RI\text{-}OR2)\text{-}H]^-$, both in CID experiments and in mass spectra suggest it is a very stable species.

To increase the signal intensity of the Aβ peptides in the Aβ:RI-OR2 mixture, and as RI-OR2 is reported to inhibit aggregation of Aβ(1-42) at substoichiometric concentrations [65], the concentration of RI-OR2 was decreased by a factor of 10 and 100. As both RI-OR2 and Aβ solutions were ziptipped before use, all concentrations are approximate. The mass spectra of the four mixtures are shown in Figure 5.11 and Figure 5.13.

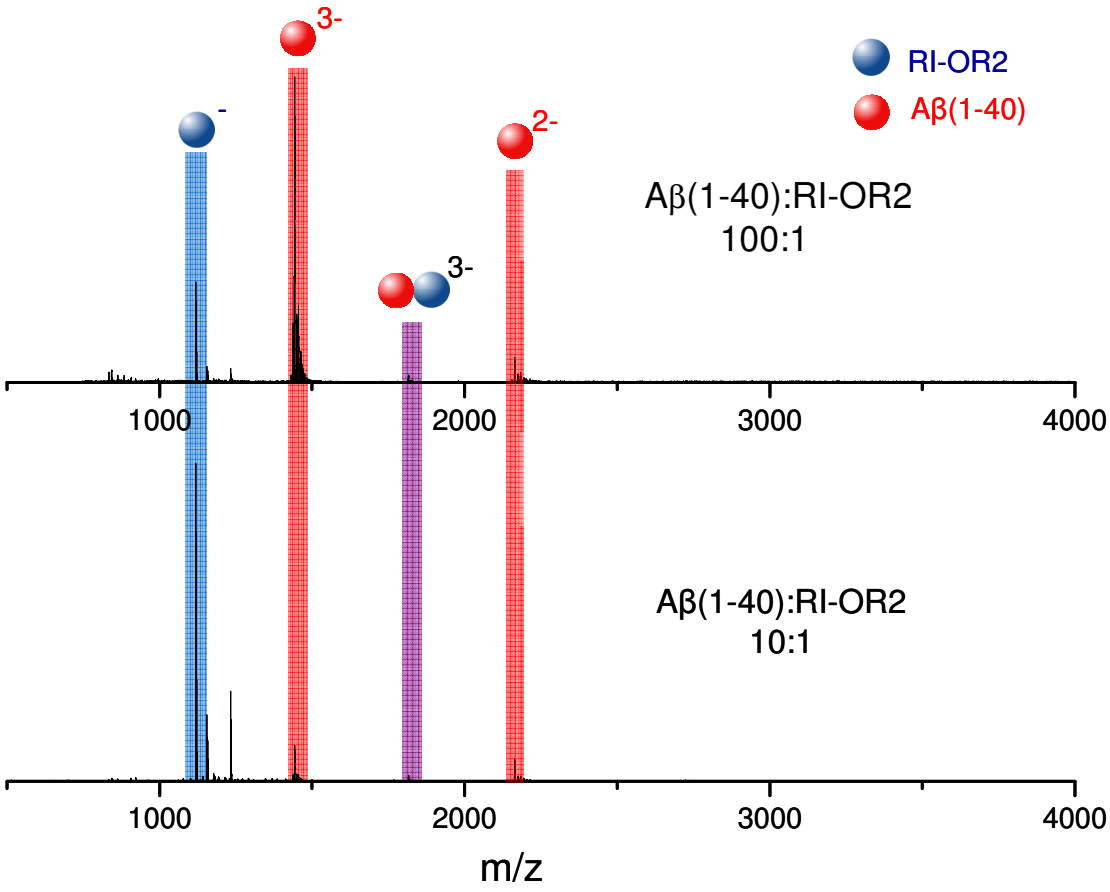


Figure 5.11 | Mass spectra of ziptipped Aβ(1-40) and ziptipped RI-OR2 in ~100:1 (50 μM Aβ(1-40): 0.5 μM RI-OR2) and ~10:1 mixture (50 μM Aβ(1-40): 5 μM RI-OR2). Solvent conditions are 49.5:49.5:1 MeOH:H_2O:HCOOH at pH 2.

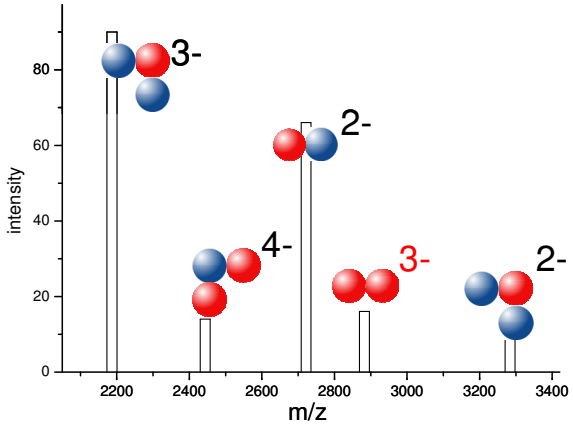


Figure 5.12 | The relative intensities of the lowest abundance complexes in ~10:1 mixture (50 μM Aβ(1-40): 5 μM RI-OR2).

In both mass spectra shown in Figure 5.11, the charge state distribution of Aβ(1-40) and RI-OR2 matches that of the peptides in isolation (Figure 5.5 and Figure 5.6). $[Aβ40+(RI\text{-}OR2)\text{-}3H]^{3-}$ is the dominant complex present both in 100:1 and 10:1 Aβ40:RI-OR2 solutions, and indeed is the only complex present in the 100:1 Aβ40:RI-OR2 solution. This suggests that the most efficient binding between Aβ(1-40) and RI-OR2 occurs when a single RI-OR2 peptide binds to an Aβ(1-40) monomer. Figure 5.12 displays the complex abundances in the 10:1 Aβ40:RI-OR2 solution, but, in brief, of the additional complexes present the next most abundant is $[Aβ40+2(RI\text{-}OR2)\text{-}3H]^{3-}$, followed by the $[Aβ40+(RI\text{-}OR2)\text{-}2H]^{2-}$ complex. The lowest abundance complexes present are $[2Aβ40+(RI\text{-}OR2)\text{-}4H]^{4-}$ (which is not observed in a 1:1 ratio Aβ40:RI-OR2 solution) and $[Aβ40+2(RI\text{-}OR2)\text{-}2H]^{2-}$ which are of a similar abundance to $[2Aβ40\text{-}3H]^{3-}$.

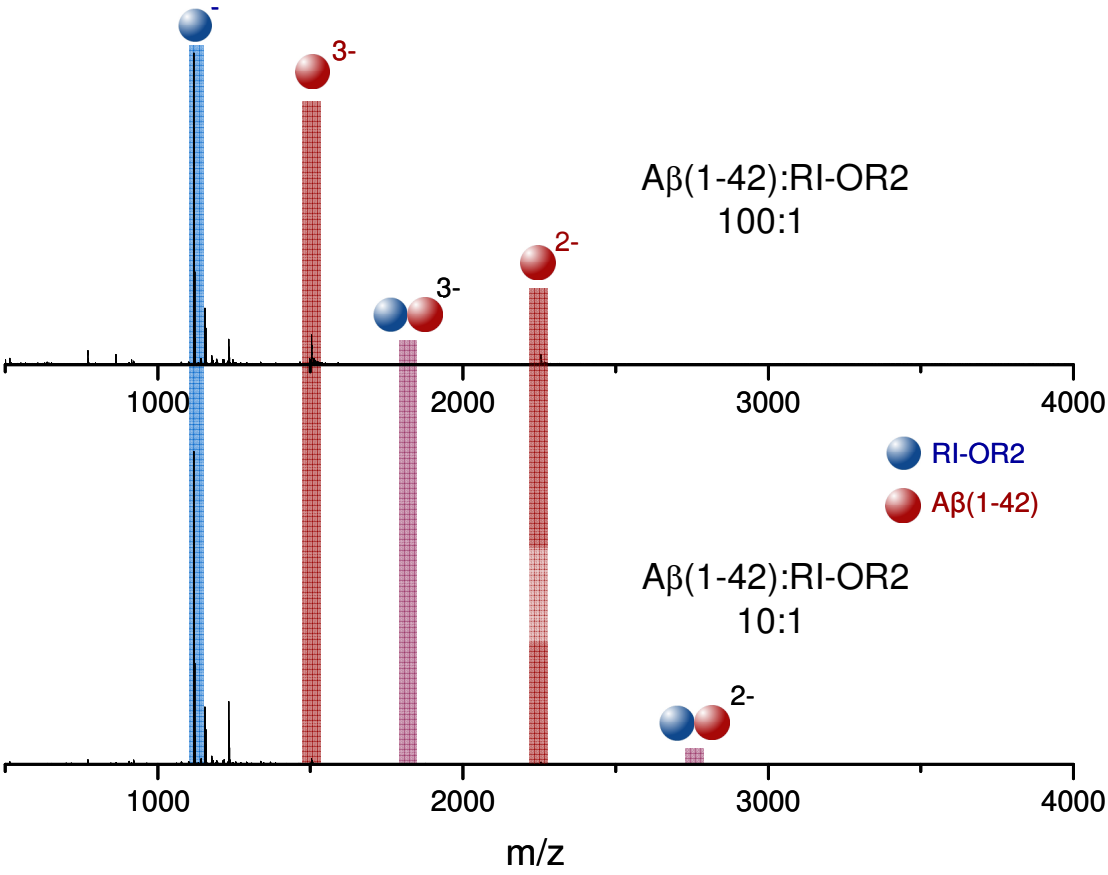


Figure 5.13 | Mass spectra of Aβ(1-42) and RI-OR2 in a ~100:1 (50 μM Aβ(1-42): 0.5 μM RI-OR2) and ~10:1 mixture (50 μM Aβ(1-42): 5 μM RI-OR2). Solvent conditions are 49.5:49.5:1 MeOH:H_2O:HCOOH at pH 2.

In both mass spectra shown in Figure 5.13, the charge state distribution of Aβ(1-42) and RI-OR2 matches that of the peptides in isolation (Figure 5.5 and Figure 5.7). $[Aβ42+(RI\text{-}OR2)-3H]^{3-}$ is the dominant complex present both in 100:1 and 10:1 Aβ42:RI-OR2 solutions. In the 10:1 Aβ42:RI-OR2 ratio solution an additional complex, $[Aβ42+(RI\text{-}OR2)-2H]^{2-}$, is observed. This suggests that, as for Aβ(1-40), the most efficient binding between of Aβ(1-42) and RI-OR2 occurs when a single RI-OR2 peptide binds to an Aβ(1-42) monomer. Only 1:1 complexes of Aβ42:RI-OR2 are observed in these solutions, this increased specificity may be why

RI-OR2 has a greater efficacy in preventing Aβ(1-42) aggregation at substoichiometric concentrations than Aβ(1-40) aggregation [65].

5.3.1.3. Positive Ionisation Mode

Positive ionisation mode MS was utilised to analyse RI-OR2 and the Aβ peptides further. Preliminary studies showed that there were far less adduction and signal suppression in positive mode than in negative. Therefore samples were not ziptipped before use.

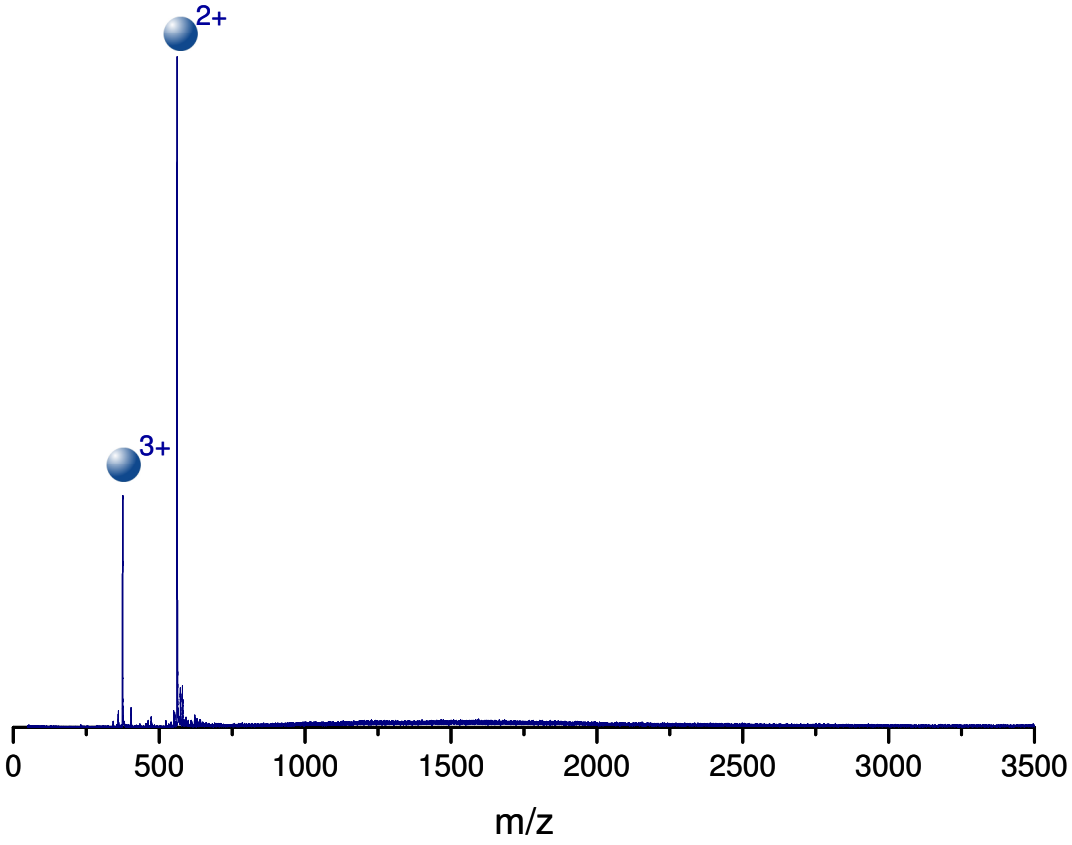


Figure 5.14 | Mass spectrum of 100 µM unziptipped RI-OR2 in aqueous HCl at pH 2.

The dominant species in the mass spectrum shown in Figure 5.14 is the $[(RI\text{-}OR2)+2H]^{2+}$ ion, with $[(RI\text{-}OR2)+3H]^{3+}$ as the only other species present. This suggests that RI-OR2 is predominantly monomeric in solution, again supporting the non-aggregating nature of RI-OR2 facilitated by the insertion of cationic arginine residues [62]. These arginine residues are the most likely sites of protonation for the dominant $[(RI\text{-}OR2)+2H]^{2+}$ species. The location of the third protonation site is likely to be the lysine residue, or, less probably, the N-terminal amine.

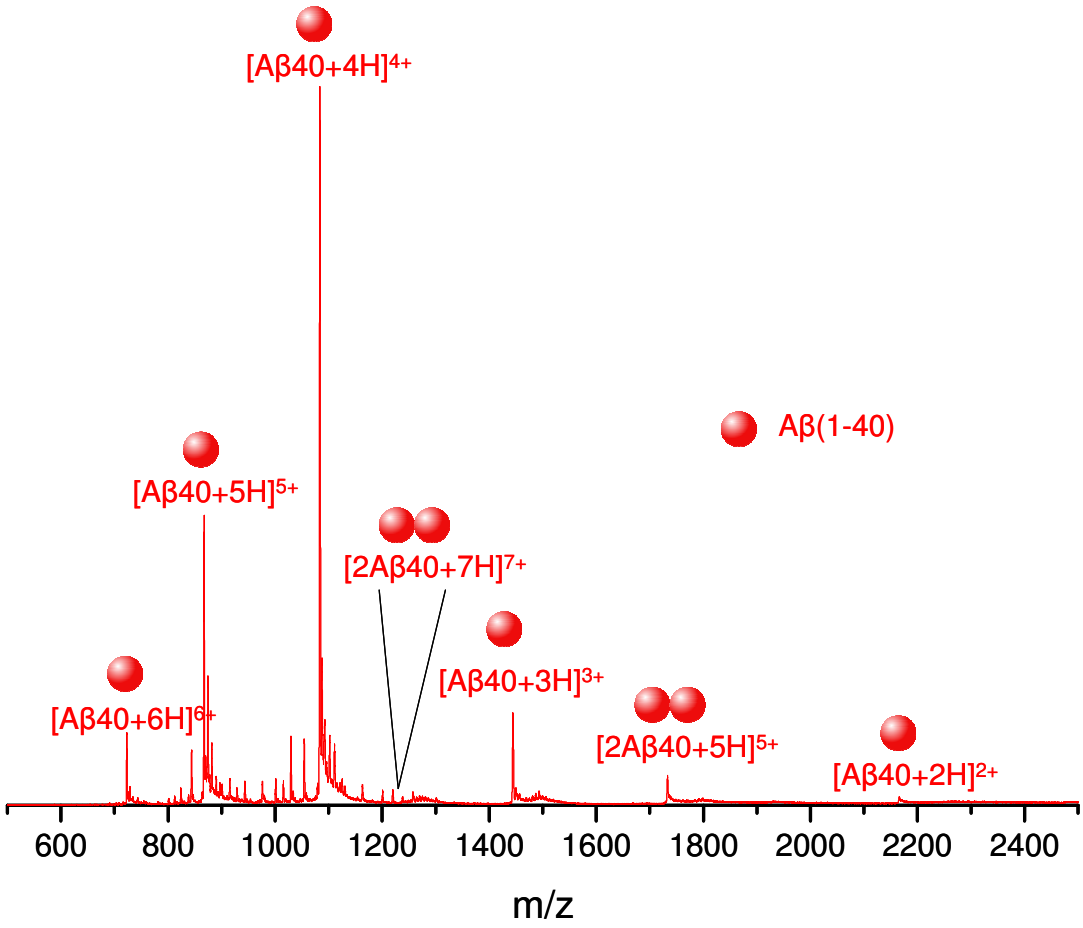


Figure 5.15 | Mass spectrum of 100 μM unziptipped Aβ(1-40) in aqueous HCl at pH 2.

$[Aβ40+4H]^{4+}$ is the dominant species present in Figure 5.15, followed by $[Aβ40+5H]^{5+}$, $[Aβ40+3H]^{3+}$ and $[Aβ40+6H]^{6+}$, suggesting Aβ(1-40) is

predominantly monomeric in solution. The most populated higher order aggregate is $[2A\beta40+5H]^{5+}$. Other oligomeric species detected in lower abundances (not visible in Figure 5.15) are $[2A\beta40+4H]^{4+}$, $[2A\beta40+6H]^{6+}$, $[3A\beta40+7H]^{7+}$, $[3A\beta40+5H]^{5+}$, $[4A\beta40+7H]^{7+}$, $[4A\beta40+9H]^{9+}$, $[5A\beta40+13H]^{13+}$ and $[6A\beta40+10H]^{10+}$. The mass spectrum acquired by Kloniecki *et al.* (supplementary information) [55] of 100 μM Aβ(1–40) in 10 mM ammonium acetate (pH 7.4) is superficially similar to Figure 5.15, as the charge state distributions of the dominant monomeric and dimeric species are alike. Table 5.2 details the Aβ(1–40) species observed in these experiments, and those reported in the literature, in both positive and negative ionisation modes.

Species	Ionisation Mode			
	Positive		Negative	
	Kloniecki *et al.* [55]	Observed in these experiments	Bernstein *et al.* [52]	Observed in these experiments
$[A\beta40+7H]^{7+}$	+	-	-	-
$[A\beta40+6H]^{6+}$	+	+	-	-
$[A\beta40+5H]^{5+}$	+	+	-	-
$[A\beta40\pm4H]^{4\pm}$	+	+	+	+
$[A\beta40\pm3H]^{3\pm}$	+	+	+	+
$[A\beta40\pm2H]^{2\pm}$	+	+	+	+
$[2A\beta40\pm7H]^{7\pm}$	+	+	-	+
$[2A\beta40+6H]^{6+}$	-	+	-	-
$[2A\beta40\pm5H]^{5\pm}$	+	+	+	-
$[2A\beta40\pm4H]^{4\pm}$	-	+	+	-
$[2A\beta40\pm3H]^{3\pm}$	+	-	-	+
$[3A\beta40+10H]^{10+}$	+	-	-	-
$[3A\beta40\pm8H]^{8\pm}$	+	-	-	+
$[3A\beta40\pm7H]^{7\pm}$	+	+	-	+
$[3A\beta40+5H]^{5+}$	+	+	-	-
$[3A\beta40+4H]^{4+}$	+	-	-	-

$[4A\beta40\text{-}10H]^{10-}$	-	-	+	-
$[4A\beta40+9H]^{9+}$	+	+	-	-
$[4A\beta40+7H]^{7+}$	+	+	-	-
$[4A\beta40\text{-}6H]^{6-}$	-	-	-	+
$[4A\beta40+5H]^{5+}$	+	-	-	-
$[5A\beta40+13H]^{13+}$	-	+	-	-
$[5A\beta40+11H]^{11+}$	+	-	-	-
$[5A\beta40+9H]^{9+}$	+	-	-	-
$[5A\beta40+8H]^{8+}$	+	-	-	-
$[5A\beta40+7H]^{7+}$	+	-	-	-
$[5A\beta40+6H]^{6+}$	+	-	-	-
$[6A\beta40+13H]^{13+}$	+	-	-	-
$[6A\beta40+11H]^{11+}$	+	-	-	-
$[6A\beta40+10H]^{10+}$	-	+	-	-
$[6A\beta40+7H]^{7+}$	+	-	-	-

Table 5.2 | Aβ(1–40) species observed in these experiments and reported in the literature in both positive and negative ionisation modes. Kloniecki et al.[55] observe many additional higher order oligomers, the largest $[16A\beta40+19H]^{19+}$, however those over hexameric order are not included in Table 5.2.

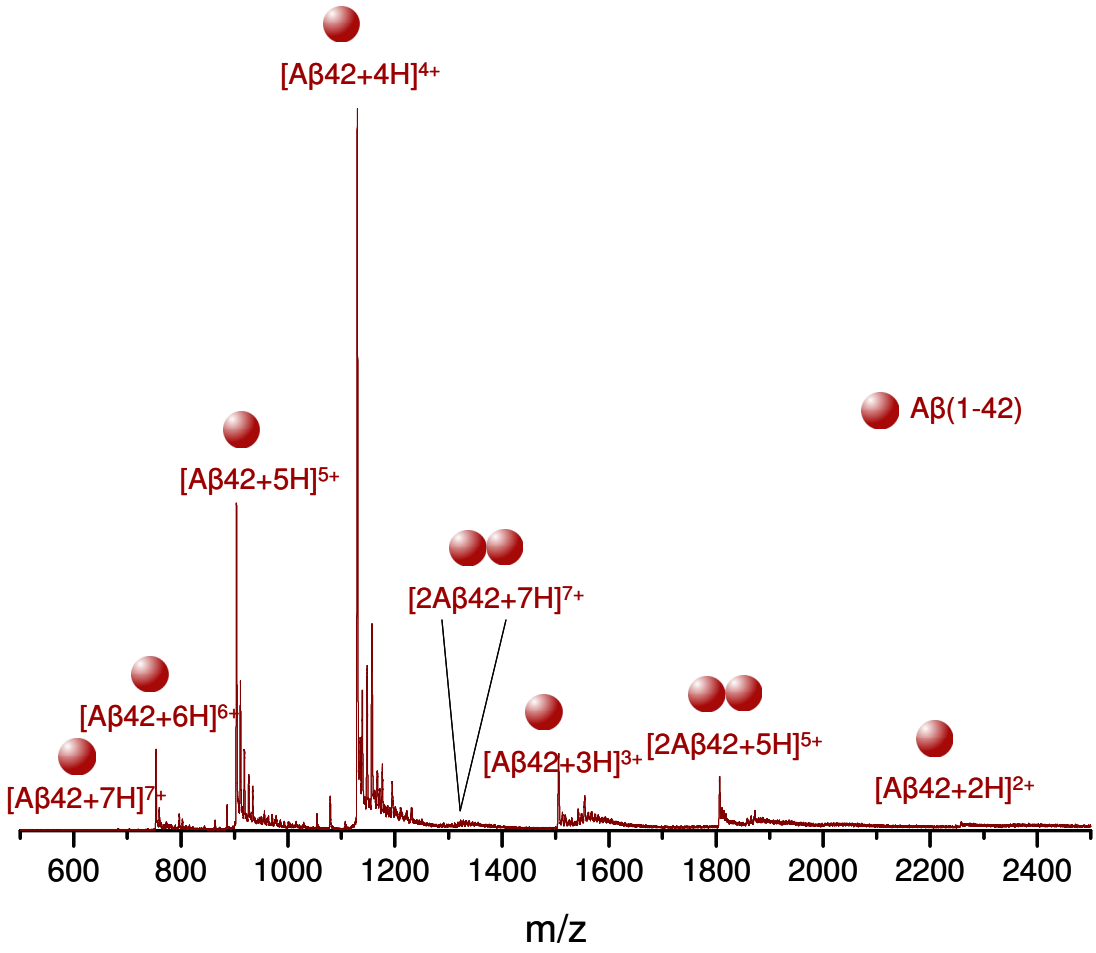


Figure 5.16 | Mass spectrum of 100 μM unziptipped Aβ(1-42) in aqueous HCl at pH 2.

$[A\beta42+4H]^{4+}$ is the dominant species in Figure 5.16, followed by $[A\beta42+5H]^{5+}$, $[A\beta42+3H]^{3+}$, $[A\beta42+6H]^{6+}$ and $[A\beta42+2H]^{2+}$. The most populated higher order aggregate is $[2A\beta42+5H]^{5+}$. Additional oligomeric species discerned, but too low in abundance to be visible in Figure 5.16, are $[A\beta42+7H]^{7+}$, $[2A\beta42+4H]^{4+}$, $[2A\beta42+6H]^{6+}$, $[2A\beta42+7H]^{7+}$, $[3A\beta42+6H]^{6+}$ and $[4A\beta42+9H]^{9+}$. The monomeric and dimeric charge state distribution and oligomeric species present are similar to that of Aβ(1-40). Bernstein *et al.* (in supplementary information) [52] observe a dominant $[A\beta42+4H]^{4+}$, a less intense $[A\beta42+3H]^{3+}$ species and a very broad *m/z* peak in the spectrum, which CID experiments suggest consists of unresolved higher

order oligomers. No experimental conditions are specified, but if they are identical to those utilised in the main body of the experimental work they are 30 μM Aβ(1-42) at pH 7.4. Table 5.3 details the Aβ(1–42) species observed in these experiments, and those reported in the literature, in both positive and negative ionisation modes.

Species	Ionisation Mode			
	Positive		Negative	
	Bernstein *et al.* [52]	Observed in these experiments	Bernstein *et al.* [51, 52]	Observed in these experiments
$[A\beta42+7H]^{7+}$	-	+	-	-
$[A\beta42+6H]^{6+}$	+	+	-	-
$[A\beta42\pm5H]^{5\pm}$	+	+	-	-
$[A\beta42\pm4H]^{4\pm}$	+	+	+	-
$[A\beta42\pm3H]^{3\pm}$	+	+	+	+
$[A\beta42\pm2H]^{2\pm}$	-	+	+	+
$[2A\beta42+7H]^{7+}$	-	+	-	-
$[2A\beta42+6H]^{6+}$	-	+	-	-
$[2A\beta42\pm5H]^{5\pm}$	-	+	+	-
$[2A\beta42+4H]^{4+}$	-	+	-	-
$[2A\beta42-3H]^{3-}$	-	-	-	+
$[3A\beta42+6H]^{6+}$	-	+	-	-
$[4A\beta42-10H]^{10-}$	-	-	+	-
$[4A\beta42+9H]^{9+}$	-	+	-	-
$[6A\beta42-15H]^{15-}$	-	-	+	-
$[10A\beta42-25H]^{25-}$	-	-	+	-
$[12A\beta42-30H]^{30-}$	-	-	+	-

Table 5.3 | Aβ(1–42) species observed in these experiments and reported in the literature in both positive and negative ionisation modes.

5.3.2. Ion Mobility Mass Spectrometry

All rotationally averaged CCSs obtained are tabulated in Appendix 5.

5.3.2.1. CCSs of Aβ(1-40), Aβ(1-42) and RI-OR2

IM-MS data were acquired for Aβ(1-40), Aβ(1-42) and RI-OR2 individually and in mixtures of peptide and inhibitor. Figure 5.17 displays the monomeric CCSs obtained in these experiments, in addition to those reported in the literature.

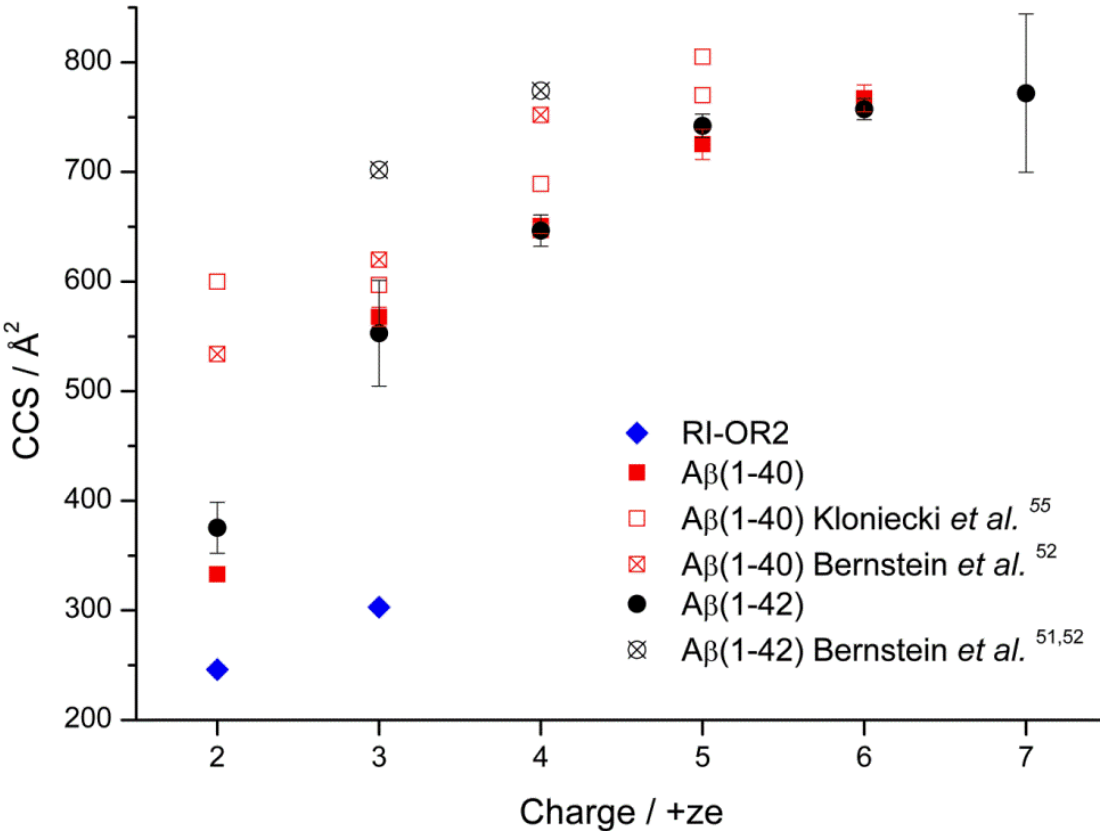


Figure 5.17 | Aβ(1-40), Aβ(1-42) and RI-OR2 monomer CCS distributions.

Figure 5.17 displays a correlation between charge state and CCS for both Aβ peptides and RI-OR2, with increasing charge giving rise to larger CCSs; a result of

stronger Coulombic repulsion. There is a large increase in CCS in our data between $[A\beta+2H]^{2+}$ and $[A\beta+3H]^{3+}$ for both Aβ(1-40) and Aβ(1-42), perhaps indicating that the addition of one charge causes unfolding of the conformational family which corresponds to $[A\beta+2H]^{2+}$. However, this trend is not present in the CCSs acquired by Bernstein *et al.* [52] where there is only a small increase in line with that observed between other charge states. The data acquired by Kloniecki *et al.* [55] displays no difference in CCS between $[A\beta40+2H]^{2+}$ and $[A\beta40+3H]^{3+}$. In our experiments there is a lesser increase in CCS with subsequent charge addition, indeed the CCSs of $[A\beta42+5H]^{5+}$, $[A\beta42+6H]^{6+}$ and $[A\beta42+7H]^{7+}$ agree within experimental error. This suggests that the extended $[A\beta42+5H]^{5+}$ conformation of Aβ(1-42) can either accommodate an additional two charges without further unfolding or that it is fully unfolded. With the exception of $[A\beta+2H]^{2+}$, where $[A\beta42+2H]^{2+}$ is significantly larger than $[A\beta40+2H]^{2+}$, monomeric CCSs of Aβ(1-40) and Aβ(1-42) agree within experimental error for the same charge states. CCS data acquired by Bernstein *et al.* [51,52] show larger Aβ(1-42) monomeric conformations than Aβ(1-40) conformations of the same charge state, however, as no standard deviation is recorded it is not possible to tell whether this difference is statistically significant. All experimental CCSs for both Aβ(1-40) and Aβ(1-42) were substantially more compact than those reported in the literature and a wider array of charge states was observed.

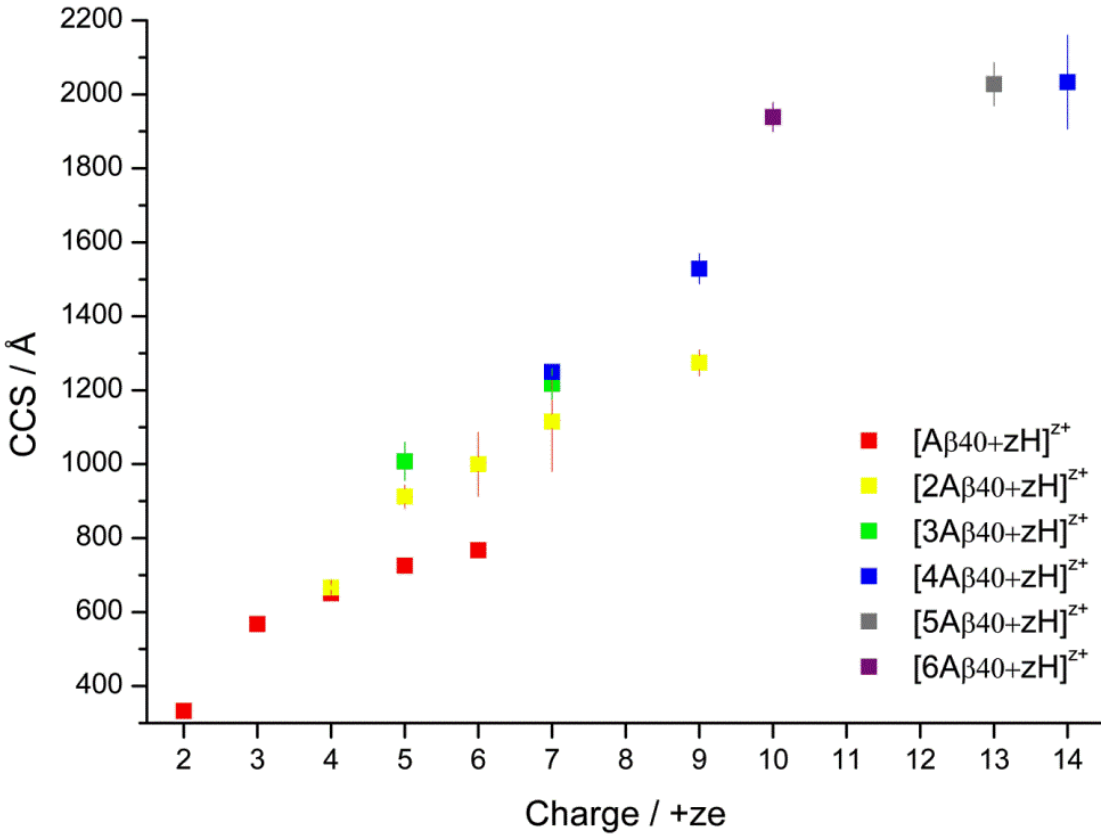


Figure 5.18 | Aβ(1-40) oligomer CCS distribution in positive ionisation mode. Data shown are an average of three repeats and error bars reflect the associated standard deviation.

In addition to the oligomeric species observed by mass spectrometry, ATDs of Aβ(1-40) species $[2A\beta40+7H]^{7+}$ and $[3A\beta40+5H]^{5+}$ revealed earlier arriving peaks which have been assigned as coincident higher order oligomers $[4A\beta40+14H]^{14+}$ and $[6A\beta40+10H]^{10+}$. Figure 5.18 displays the CCSs of all Aβ(1-40) oligomers observed as a function of charge. The expected increase in CCS with charge and multimeric order is observed, however, $[2A\beta40+7H]^{7+}$, $[3A\beta40+7H]^{7+}$ and $[4A\beta40+7H]^{7+}$ agree within experimental standard deviation, suggesting a packing phenomenon akin to that seen for TTR(105-115) (Chapter Three, Figure 3.11). The highest charge states of the tetramer, pentamer and hexamer species have CCSs which agree within experimental error.

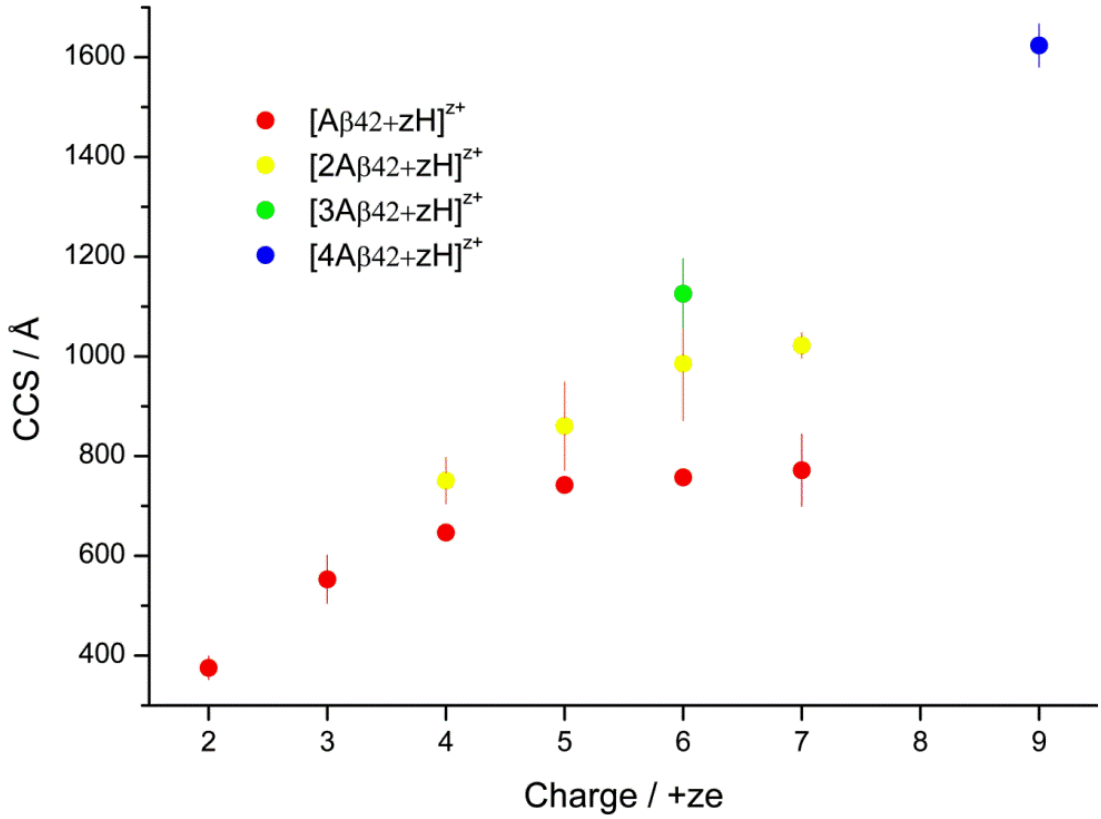


Figure 5.19 | Aβ(1-42) CCS distribution in positive ionisation mode. Data shown are an average of three repeats and error bars reflect the associated standard deviation.

Figure 5.19 shows the CCSs of Aβ(1-42) oligomers as a function of charge. Aβ(1-42) oligomers of up to tetrameric order are observed, and for most there is a correlation between increasing oligomeric order, charge and CCS. The CCSs of all Aβ(1-42) dimers agree, within experimental standard deviation, to that of a dimer of adjacent charge state. However, there is an increase in the mean values with charge. $[2A\beta42+6H]^{6+}$ and $[3A\beta42+6H]^{6+}$ agree within experimental error, echoing the packing tendency described in more detail previously.

5.3.2.2. Mass Spectra and CCSs of Complexes between Aβ Peptides and RI-OR2

5.3.2.2.1. Aβ(1-40) and RI-OR2

Mass spectra of unziptipped Aβ(1-40) and RI-OR2 at a 1:1 ratio, Figure 5.20, revealed charge state distributions similar to those present when the peptides were analysed separately (Figure 5.14 and Figure 5.15). Seven complexes of Aβ(1-40) and RI-OR2 were detected: $[Aβ40+(RI-OR2)+3H]^{3+}$, $[Aβ40+(RI-OR2)+4H]^{4+}$, $[Aβ40+(RI-OR2)+5H]^{5+}$, $[Aβ40+(RI-OR2)+6H]^{6+}$, $[Aβ40+2(RI-OR2)+3H]^{3+}$, $[Aβ40+2(RI-OR2)+4H]^{4+}$ and $[2Aβ40+(RI-OR2)+5H]^{5+}$, although only the most abundant complexes, $[Aβ40+(RI-OR2)+4H]^{4+}$ and $[2Aβ40+(RI-OR2)+5H]^{5+}$, are visible in Figure 5.20.

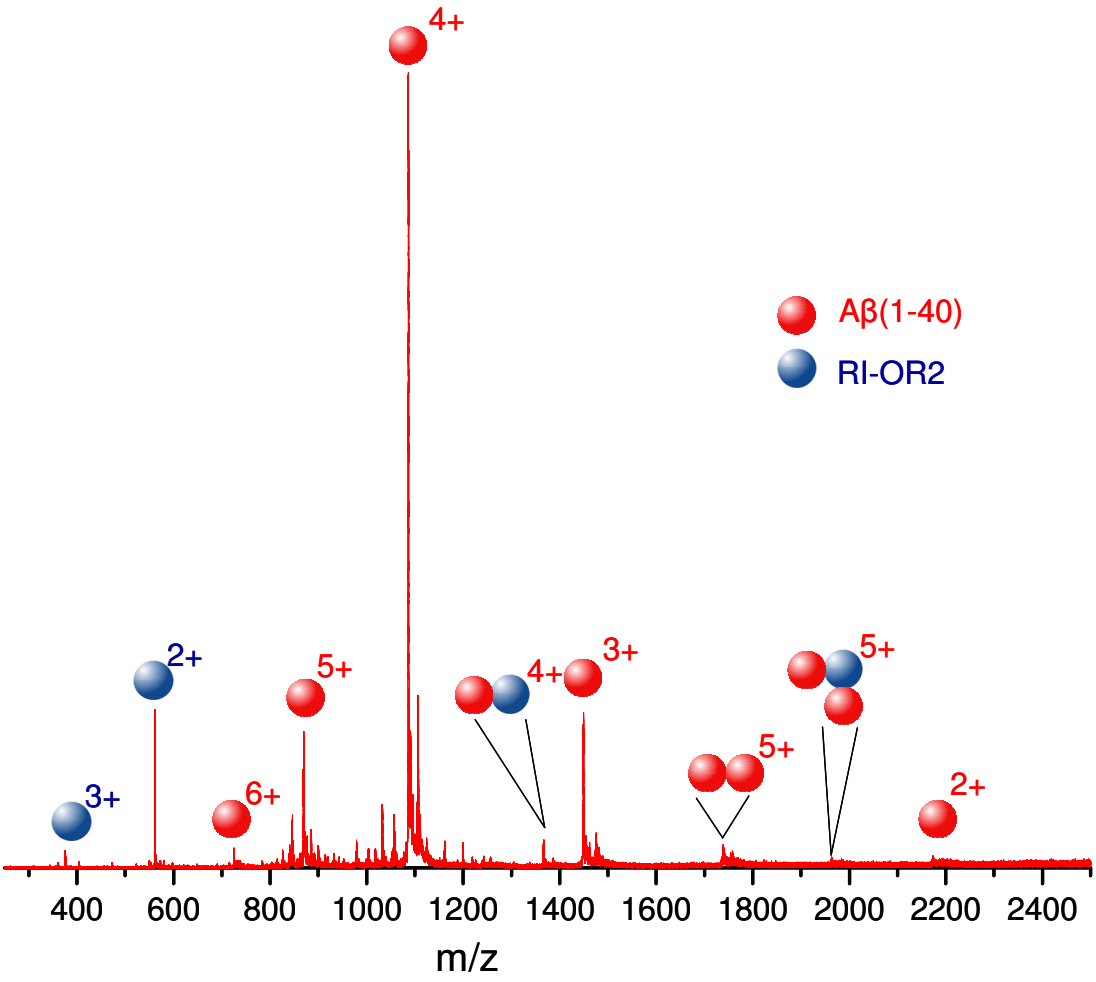


Figure 5.20 | Mass spectrum of unziptipped Aβ(1-40) and RI-OR2 at a 1:1 ratio and concentrations of 50 μM in pH 2 aqueous HCl.

Rotationally averaged CCSs were calculated for all species and complexes observed in the mass spectra and are displayed in Figure 5.21. For clarity Figure 5.21 does not display the CCSs of Aβ(1-40) oligomers larger than the trimeric species; CCSs of all Aβ(1-40) oligomers can be found in Figure 5.18 and tabulated in Appendix 5.

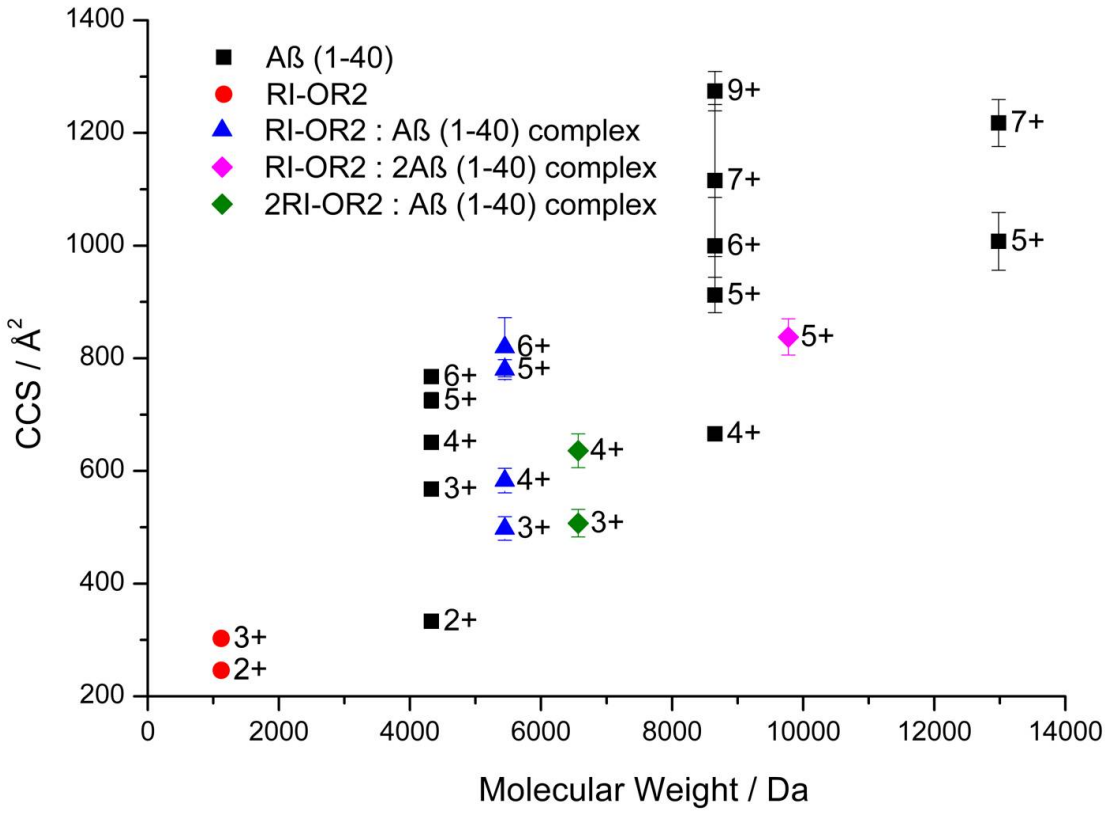


Figure 5.21 | The CCSs of Aβ(1-40) up to trimeric oligomeric order, RI-OR2 and complexes of Aβ(1-40) and RI-OR2. Aβ(1-40) and RI-OR2 were unziptipped at concentrations of 50 μM in pH 2 aqueous HCl and a 1:1 ratio. Data shown are an average of three repeats and error bars reflect the associated standard deviation.

Of the seven Aβ(1-40):RI-OR2 complexes observed, the majority display a conformational compaction; $[Aβ40+(RI-OR2)+3H]^{3+}$, $[Aβ40+(RI-OR2)+4H]^{4+}$ and $[2Aβ40+(RI-OR2)+5H]^{5+}$ all are appreciably smaller than their counterpart species $[Aβ40+3H]^{3+}$, $[Aβ40+4H]^{4+}$ and $[2Aβ40+5H]^{5+}$. This is unexpected as each complex is ~1.1 kDa more massive than its counterpart Aβ(1-40) species and such a

significant increase in mass would be expected to result in an appreciable increase in CCS. $[Aβ40+2(RI-OR2)+3H]^{3+}$ has a CCS which agrees within experimental standard deviation with $[Aβ40+(RI-OR2)+3H]^{3+}$, which is itself significantly more compact than $[Aβ40+3H]^{3+}$. The CCSs of $[Aβ40+2(RI-OR2)+4H]^{4+}$ agrees within experimental standard deviation with $[Aβ40+4H]^{4+}$, despite the ~2.2 kDa disparity in mass between the two species. A possible mechanism for the mode of action of RI-OR2 could be a compaction of Aβ(1-40) monomers in solution, thus preventing the presence of an elongated, aggregation competent monomeric species.

However, there are complexes which are less compact than their corresponding Aβ(1-40) monomers. $[Aβ40+(RI-OR2)+5H]^{5+}$ possesses a larger CCS, and thus is in a more extended conformation, than $[Aβ40+5H]^{5+}$, whereas the CCSs of $[Aβ40+(RI-OR2)+6H]^{6+}$ and $[Aβ40+6H]^{6+}$ agree within experimental standard deviations. As in the mass spectrum $[Aβ40+6H]^{6+}$ and $[Aβ40+5H]^{5+}$ are the least intense monomeric species present, it can be inferred that they are less populated in solution than other Aβ(1-40) monomeric species. These extended monomeric conformations, and the lack of compaction RI-OR2 binding entails, could be the cause of the lesser efficacy of RI-OR2 in preventing Aβ(1-40) aggregation [65].

5.3.2.2.2. Aβ(1-42) and RI-OR2

From Figure 5.22 it can be observed that the charge state distributions of Aβ(1-42) and RI-OR2 are very similar to those when the two peptides are individually analysed (Figure 5.14 and Figure 5.16). Three complexes of Aβ(1-42) and RI-OR2 are detected; the most populated is $[Aβ42+(RI-OR2)+4H]^{4+}$ (this is the only complex shown in Figure 5.22), followed by $[Aβ42+(RI-OR2)+3H]^{3+}$ and finally $[Aβ42+2(RI-OR2)+4H]^{4+}$.

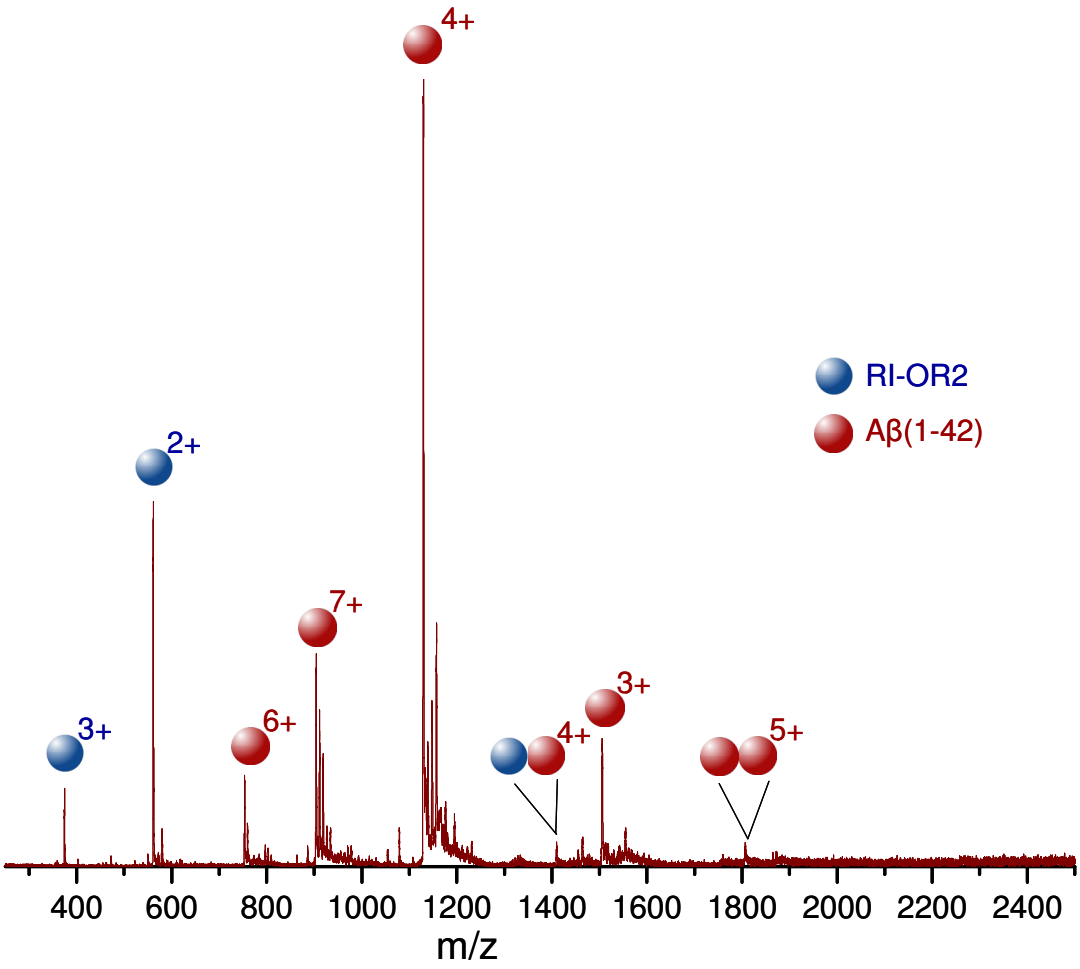


Figure 5.22 | Mass spectrum of unziptipped Aβ(1-42) and RI-OR2 at an approximately 1:1 ratio and concentrations of 50 μM in pH 2 aqueous HCl.

Rotationally averaged CCSs were calculated for all species observed in the mass spectra. For clarity, Figure 5.23 does not display the CCSs of Aβ(1-42) oligomers larger than the trimeric species; CCSs of all Aβ(1-42) oligomers observed are displayed in Figure 5.19 and tabulated in Appendix 5.

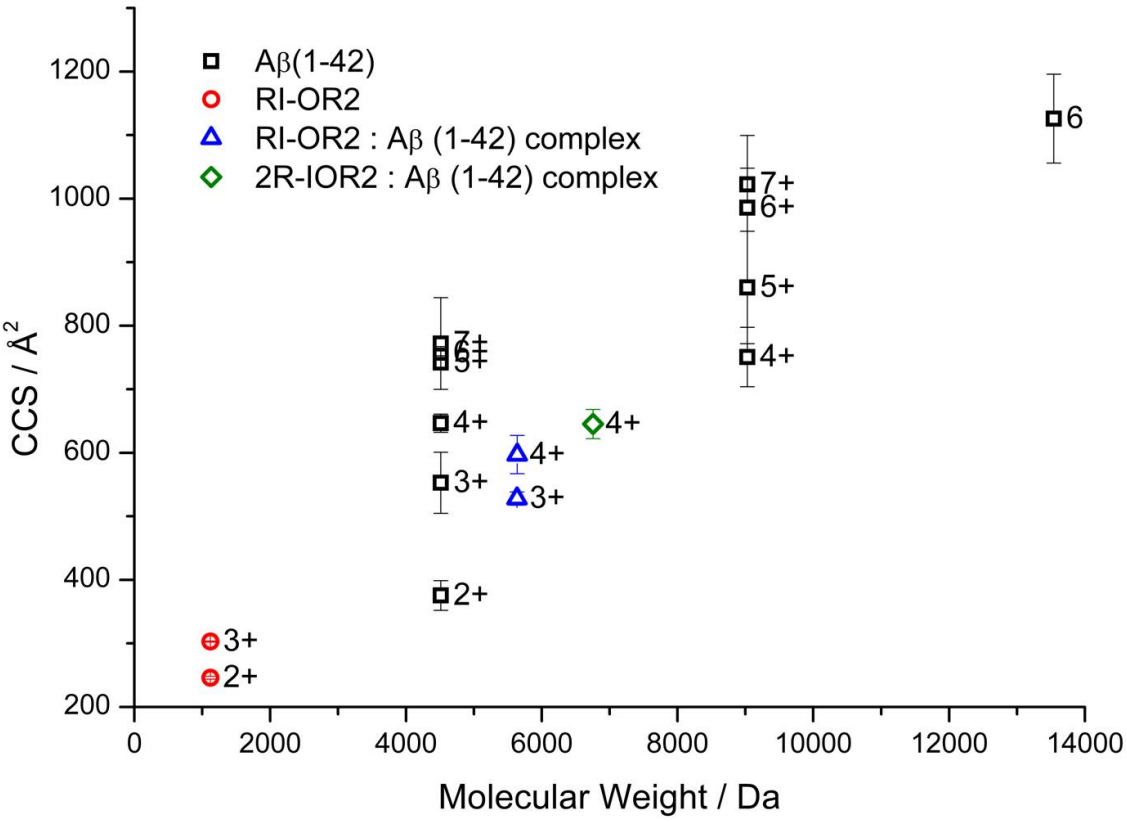


Figure 5.23 | The CCSs of Aβ (1-42), RI-OR2 and complexes up to trimeric oligomeric order. Aβ(1-42) and RI-OR2 were unziptipped at concentrations of 50 μM in pH 2 aqueous HCl and a 1:1 ratio. Data shown are an average of three repeats (except [Aβ42+(RI-OR2)+3H]$^{3+}$ which is an average of two repeats) , and error bars reflect the associated standard deviation.

The CCS of [Aβ42+(RI-OR2)+4H]$^{4+}$ is significantly more compact than [Aβ42+4H]$^{4+}$, but agrees to within experimental error with [Aβ42+3H]$^{3+}$. The complex with an additional RI-OR2 monomer, [Aβ42+2(RI-OR2)+4H]$^{4+}$, has a CCS which agrees within experimental standard deviation with the ~2.2 kDa less massive [Aβ40+4H]$^{4+}$. Although the mean CCS of [Aβ42+(RI-OR2)+3H]$^{3+}$ is less than that of [Aβ42+3H]$^{3+}$, within experimental error their CCSs agree.

The mechanism by which RI-OR2 inhibits Aβ oligomer formation could be by sequestering the monomers into a more compact structure which, in the case of Aβ(1-42), could prevent their association into an 'open tetramer' and from there to a paranucleus [51]. The presence of more extended Aβ(1-40) conformers complexed with RI-OR2 could be the reason why RI-OR2 is a less effective inhibitor of Aβ(1-40) aggregation. This mechanism is depicted in Figure 5.24.

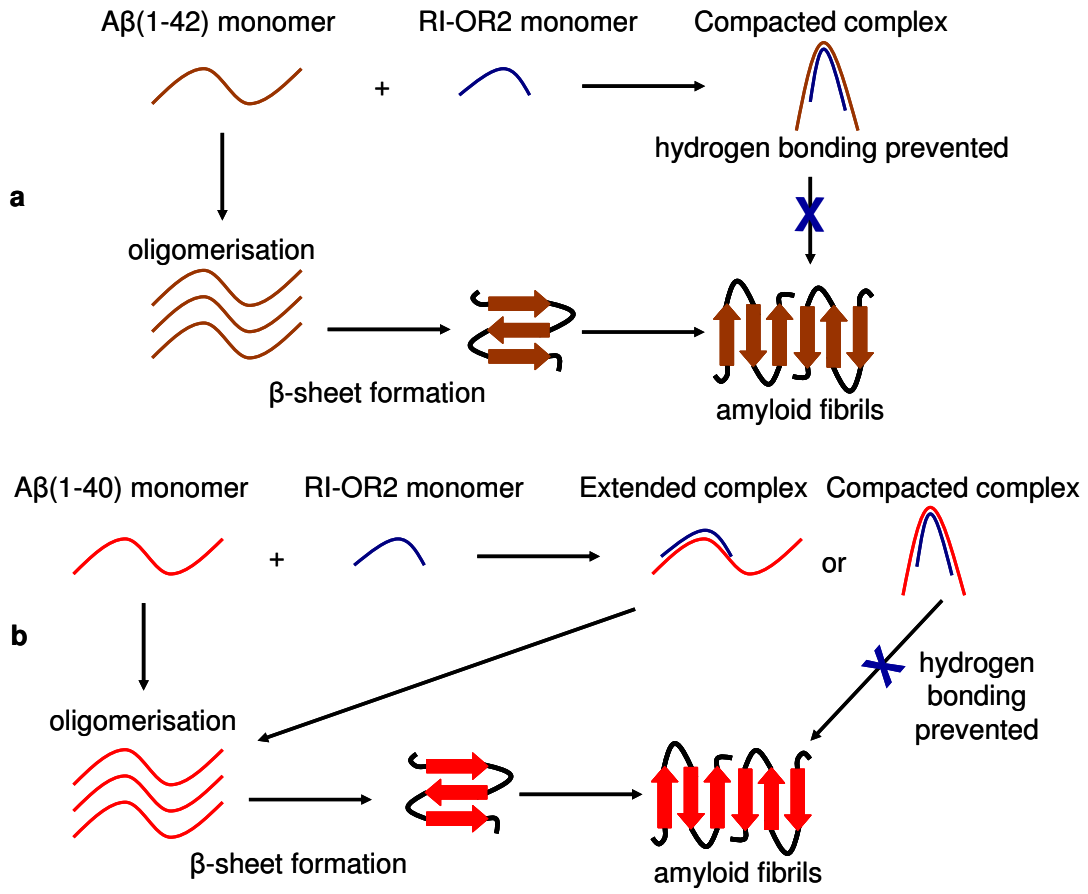


Figure 5.24 | Postulated mechanisation of RI-OR2 action to inhibit the oligomerisation and fibril formation a) Aβ(1-42) b) Aβ(1-40).

5.4. Conclusions

Mass spectra of RI-OR2 in negative and positive ionisation modes reveal predominantly monomeric states. The dominant species of Aβ(1-40) and Aβ(1-42) are monomers, in both ionisation modes. Aβ(1-40) oligomers were observed of up to tetrameric order in negative ionisation mode and hexameric order in positive ionisation mode. Fewer higher order oligomers were observed for Aβ(1-42): dimers in negative ionisation mode and a tetrameric species in positive ionisation mode. Complexes of Aβ and RI-OR2 were observed when the Aβ peptides were mixed with the inhibitor. Complex stoichiometries of Aβ:RI-OR2, Aβ:2RI-OR2 and 2Aβ:RI-OR2 are observed but 1:1 complexes are generally more populated than other stoichiometries. In the majority of cases complexes of Aβ and RI-OR2 possess substantially more compact CCSs than the corresponding Aβ species of the same charge state. This compaction of an elongated Aβ species could be the mechanism by which RI-OR2 inhibits Aβ oligomer formation. Further investigations could be fruitful in characterising interactions between RI-OR2 and Aβ peptides, providing insight into the mode of action of this potential therapeutic.

5.5. References

1 Tanzi, R. E. *et al.* Amyloid Beta-Protein Gene - CDNA, Messenger-RNA Distribution, and Genetic-Linkage near the Alzheimer Locus. *Science* **235**, 880-884 (1987).

2 Sabo, S. L., Ikin, A. F., Buxbaum, J. D. & Greengard, P. The Alzheimer amyloid precursor protein (APP) and FE65, an APP-binding protein, regulate cell movement. *J. Cell Biol.* **153**, 1403-1414 (2001).

3 Glenner, G. G. & Wong, C. W. Alzheimers Disease - Initial Report of the Purification and Characterisation of a Novel Cerebrovascular Amyloid Protein. *Biochem. Biophys. Res. Commun.* **120**, 885-890 (1984).

4 Jarrett, J. T., Berger, E. P. & Lansbury, P. T. The Carboxy Terminus of the β-Amyloid Protein is Critical for the Seeding Of Amyloid Formation - Implications for the Pathogenesis of Alzheimers Disease. *Biochemistry* **32**, 4693-4697 (1993).

5 Kojro, E. & Fahrenholz, F. in *Subcellular Biochemistry* **38,** 105-127 (Springer, New York, United States, 2005).

6 Hardy, J. & Selkoe, D. J. Medicine - The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353-356 (2002).

7 Hartmann, T. *et al.* Distinct sites of intracellular production for Alzheimer's disease Aβ 40/42 amyloid peptides. *Nat. Med.* **3**, 1016-1020 (1997).

8 Roher, A. E. *et al.* β-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *P. Natl. Acad. Sci. U. S. A.* **90**, 10836-10840 (1993).

9 Mayeux, R. *et al.* Plasma amyloid β-peptide 1–42 and incipient Alzheimer's disease. *Ann. Neurol.* **46**, 412-416 (1999).

10 Klein, W. L., Krafft, G. A. & Finch, C. E. Targeting small Aβ oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci.* **24**, 219-224 (2001).

11 Younkin, S. G. Evidence that Aβ-42 is the Real Culprit in Alzheimers-Disease. *Ann. Neurol.* **37**, 287-288 (1995).

12 Selkoe, D. J. Translating cell biology into therapeutic advances in Alzheimer's disease. *Nature* **399**, A23-A31 (1999).

13 Dahlgren, K. N. *et al.* Oligomeric and fibrillar species of amyloid-β peptides differentially affect neuronal viability. *J. Biol. Chem.* **277** (2002).

14 Jarrett, J. T., Berger, E. P. & Lansbury, P. T., Jr. The carboxy terminus of the β amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32**, 4693-4697 (1993).

15 Kirschner, D. A., Abraham, C. & Selkoe, D. J. X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-β conformation. *P. Natl. Acad. Sci. U. S. A.* **83**, 503-507 (1986).

16 Ahmed, M. *et al.* Structural conversion of neurotoxic amyloid-β 1-42 oligomers to fibrils. *Nat. Struct. Mol. Biol.* **17**, 561-567 (2010).
17 Luhrs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1-42) fibrils. *P. Natl. Acad. Sci. U. S. A.* **102**, 17342-17347 (2005).
18 Masuda, Y. *et al.* Verification of the C-terminal intramolecular β-sheet in Aβ42 aggregates using solid-state NMR: implications for potent neurotoxicity through the formation of radicals. *Bioorg. Med. Chem. Lett.* **18**, 3206-3210 (2008).
19 Tycko, R. Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q. Rev. Biophys.* **39**, 1-55 (2006).
20 Paravastu, A. K., Leapman, R. D., Yau, W. M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *P. Natl. Acad. Sci. U. S. A.* **105**, 18349-18354 (2008).
21 Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *P. Natl. Acad. Sci. U. S. A.* **99**, 16742-16747 (2002).
22 Antzutkin, O. N. *et al.* Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of β-sheets in Alzheimer's β-amyloid fibrils. *P. Natl. Acad. Sci. U. S. A.* **97**, 13045-13050 (2000).
23 Balbach, J. J. *et al.* Supramolecular structure in full-length Alzheimer's β-amyloid fibrils: Evidence for a parallel β-sheet organization from solid-state nuclear magnetic resonance. *Biophys. J.* **83**, 1205-1216 (2002).
24 Kheterpal, I., Zhou, S., Cook, K. D. & Wetzel, R. Aβ amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *P. Natl. Acad. Sci. U. S. A.* **97**, 13597-13601 (2000).
25 Kheterpal, I., Williams, A., Murphy, C., Bledsoe, B. & Wetzel, R. Structural features of the Aβ amyloid fibril elucidated by limited proteolysis. *Biochemistry* **40**, 11757-11767 (2001).
26 Williams, A. D. *et al.* Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335**, 833-842 (2004).
27 Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307**, 262-265 (2005).
28 Olofsson, A., Sauer-Eriksson, A. E. & Ohman, A. The solvent protection of alzheimer amyloid-β-(1-42) fibrils as determined by solution NMR spectroscopy. *J. Biol. Chem.* **281**, 477-483 (2006).
29 Torok, M. *et al.* Structural and Dynamic Features of Alzheimer's Aβ Peptide in Amyloid Fibrils Studied by Site-directed Spin Labeling. *J. Biol. Chem.* **277**, 40810-40815 (2002).
30 Antzutkin, O. N., Leapman, R. D., Balbach, J. J. & Tycko, R. Supramolecular structural constraints on Alzheimer's β-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* **41**, 15436-15450 (2002).
31 Harper, J. D., Lieber, C. M. & Lansbury, P. T. Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-β protein. *Chem. Biol.* **4**, 951-959 (1997).
32 Harper, J. D., Wong, S. S., Lieber, C. M. & Lansbury, P. T. Observation of metastable Aβ amyloid protofibrils by atomic force microscopy. *Chem. Biol.* **4**, 119-125 (1997).

33 Walsh, D. M., Lomakin, A., Benedek, G. B., Condron, M. M. & Teplow, D. B. Amyloid β-protein fibrillogenesis - Detection of a protofibrillar intermediate. *J. Biol. Chem.* **272**, 22364-22372 (1997).

34 Walsh, D. M. *et al.* Amyloid β-protein fibrillogenesis - Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **274**, 25945-25952 (1999).

35 Williams, A. D. *et al.* Structural properties of Aβ protofibrils stabilized by a small molecule. *P. Natl. Acad. Sci. U. S. A.* **102**, 7115-7120 (2005).

36 Grasso, G. The use of mass spectrometry to study amyloid-β peptides. *Mass Spectrom. Rev.* **30**, 347-365 (2010).

37 Bitan, G. *et al.* Amyloid beta-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *P. Natl. Acad. Sci. U. S. A.* **100**, 330-335 (2003).

38 Bitan, G., Lomakin, A. & Teplow, D. B. Amyloid β-protein oligomerization - Prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *J. Biol. Chem.* **276**, 35176-35184 (2001).

39 Roher, A. E. *et al.* Morphology and toxicity of Aβ-(1-42) dimer derived from neuritic and vascular amyloid deposits of Alzheimer's disease. *J. Biol. Chem.* **271**, 20631-20635 (1996).

40 Podlisny, M. B. *et al.* Aggregation of Secreted Amyloid β-Protein into Sodium Dodecyl Sulfate-Stable Oligomers in Cell-Culture. *J. Biol. Chem.* **270**, 9564-9570 (1995).

41 Walsh, D. M., Tseng, B. P., Rydel, R. E., Podlisny, M. B. & Selkoe, D. J. The oligomerization of amyloid β-protein begins intracellularly in cells derived from human brain. *Biochemistry* **39**, 10831-10839 (2000).

42 Lesne, S. *et al.* A specific amyloid-β protein assembly in the brain impairs memory. *Nature* **440**, 352-357 (2006).

43 Mastrangelo, I. A. *et al.* High-resolution atomic force microscopy of soluble Aβ 42 oligomers. *J. Mol. Biol.* **358**, 106-119 (2006).

44 Frydman-Marom, A. *et al.* Cognitive-Performance Recovery of Alzheimer's Disease Model Mice by Modulation of Early Soluble Amyloidal Assemblies. *Angew. Chem. Int. Edit.* **48**, 1981-1986 (2009).

45 Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267-298 (2003).

46 Glabe, C. G. Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiol. Aging* **27**, 570-575 (2006).

47 McLean, C. A. *et al.* Soluble pool of Aβ amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **46**, 860-866 (1999).

48 Lue, L. F. *et al.* Soluble amyloid β peptide concentration as a predictor of synaptic change in Alzheimer's disease. *Am. J. Pathol.* **155**, 853-862 (1999).

49 Hsia, A. Y. *et al.* Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *P. Natl. Acad. Sci. U. S. A.* **96**, 3228-3233 (1999).

50 Baumketner, A. *et al.* Amyloid β-protein monomer structure: A computational and experimental study. *Protein Sci.* **15**, 420-428 (2006).

51 Bernstein, S. L. *et al.* Amyloid β-protein: Monomer structure and early aggregation states of Aβ 42 and its Pro(19) alloform. *J. Am. Chem. Soc.* **127**, 2075-2084 (2005).

52 Bernstein, S. L. *et al.* Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326-331 (2009).

53 Bitan, G., Vollers, S. S. & Teplow, D. B. Elucidation of primary structure elements controlling early amyloid β-protein oligomerization. *J. Biol. Chem.* **278**, 34882-34889 (2003).

54 Murray, M. M. *et al.* Amyloid β Protein: Aβ 40 Inhibits Aβ 42 Oligomerization. *J. Am. Chem. Soc.* **131**, 6316-+ (2009).

55 Kloniecki, M. *et al.* Ion Mobility Separation Coupled with MS Detects Two Structural States of Alzheimer's Disease Aβ 1-40 Peptide Oligomers. *J. Mol. Biol.* **407**, 110-124 (2011).

56 Tjernberg, L. O. *et al.* Arrest of β-amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* **271**, 8545-8548 (1996).

57 Tjernberg, L. O. *et al.* A molecular model of Alzheimer amyloid β-peptide fibril formation. *J. Biol. Chem.* **274**, 12619-12625 (1999).

58 Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* **4**, 49-60 (2003).

59 Soto, C., Kindy, M. S., Baumann, M. & Frangione, B. Inhibition of Alzheimer's amyloidosis by peptides that prevent β-sheet conformation. *Biochem. Biophys. Res. Commun.* **226**, 672-680 (1996).

60 Soto, C. *et al.* β-sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: Implications for Alzheimer's therapy. *Nat. Med.* **4**, 822-826 (1998).

61 Necula, M., Kayed, R., Milton, S. & Glabe, C. G. Small molecule inhibitors of aggregation indicate that amyloid β oligomerization and fibrillization pathways are independent and distinct. *J. Biol. Chem.* **282**, 10311-10324 (2007).

62 Austen, B. M. *et al.* Designing peptide inhibitors for oligomerization and toxicity of Alzheimer's β-amyloid peptide. *Biochemistry* **47**, 1984-1992 (2008).

63 Chorev, M. & Goodman, M. A Dozen Years of Retro-Inverso Peptidomimetics. *Accounts Chem. Res.* **26**, 266-273 (1993).

64 Chorev, M. & Goodman, M. Recent Developments in Retro Peptides and Proteins - an Ongoing Topochemical Exploration. *Trends Biotechnol.* **13**, 438-445 (1995).

65 Taylor, M. *et al.* Development of a Proteolytically Stable Retro-Inverso Peptide Inhibitor of β-Amyloid Oligomerization as a Potential Novel Treatment for Alzheimer's Disease. *Biochemistry* **49**, 3261-3272 (2010).

66 McCullough, B. J. *et al.* Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal. Chem.* **80**, 6336-6344 (2008).

6

Conclusions

Mass spectrometry, ion mobility mass spectrometry and molecular modelling have provided valuable insights in to the early aggregation of three systems: TTR (105-115), insulin and Aβ peptides. In this chapter the conclusions of this thesis are presented and further work is suggested.

In Chapter Three, MS and IM-MS were utilised to reveal the changes of oligomer populations in an aggregating solution of the amyloidogenic endecapeptide TTR (105-115). Initially oligomeric species from dimeric to tridecameric order are observed and the relative populations of these fluctuate over an eight hour time period. Throughout the time course the dimer is the most prevalent species, others of note are a stable tetrameric species which persisted throughout the experimental time period and an unusually prevalent and transiently stable decamer. Although neither MS nor IM-MS allows the differentiation between on and off pathway species, identifying these transient oligomers by their *m/z* ratio and mobility is important as, for disease related proteins or peptides, they may play a role in pathogenesis.

IM-MS elucidated the conformational space TTR (105-115) oligomers inhabit; for oligomers which populate multiple charge states there is a correlation between increasing charge and increasing CCS, due to repulsive Coulombic forces. Although oligomer CCSs increase in a linear fashion for the highest charge state for each oligomer, there is a lack of increase in CCS when $n > 3$ for oligomers of the same charge state. A large increase in oligomer mass can result in only a small increase in CCS, suggesting dense packing. This suggests there is a competition between Coulombic repulsion and packing to form a dense aggregate. All oligomers for which $n > 2$ carry fewer charges than the number of constituent peptides and each peptide's most probable protonation site is the N-terminus. This implies some N-termini are inaccessible and packed in the interior of the aggregate and, given the amino acid sequence of TTR (105–115), it is likely that this packing is governed by hydrophobic interactions. Comparison of experimental oligomer CCSs with simulations revealed that the oligomers observed are compact globular structures and contain no extended β-sheet elements. Investigations into TTR (105-115) using double IR-UV double resonance spectroscopy, in collaboration with Professor Tom Rizzo [*], are currently being undertaken by Jakub Ujma, a member of Dr Perdita Barran's research group. Ions are electrosprayed, mass selected and cooled to ~6 K in a linear 22 pole ion trap. IR and UV laser pulses irradiate the ions and the resulting spectra from detected

[*] Laboratory of Molecular Physical Chemistry, Ecole Polytechnique Fédérale de Lausanne.

fragments produced after photon absorption are collected [†]. This will be used unequivocally confirm or refute the presence of β-sheet in TTR (105-115) oligomers.

In Chapter Four the early aggregation of bovine insulin was studied; the most populated species are monomers and dimers, but oligomers up to dodecameric order are detected before the onset of accelerated fibril growth. An experimental time course was undertaken but no significant changes in the relative oligomer populations were observed. At a zero time point in the aggregating solution, the orthogonal dimension provided by IM-MS facilitates the observation of multiple conformations of insulin oligomers from dimeric to hexameric order and the identification of higher order aggromers composed of up to 24 monomers. The presence of multiple oligomer conformations are confirmed by FT-ICR MS and CID measurements, and for all oligomeric species observed, the most significantly populated conformation is the most compact form. Injection energy experiments and the empirical de la Mora relation suggest that extended conformations are not the result of gas phase unfolding events but represent structures present in solution.

Molecular modelling was employed to characterise the multiple conformations of the prevalent 7+ dimer. Both the compact and extended experimental CCSs could be replicated and modelling suggests that the conformationally compact, most experimentally prevalent, dimeric species is enriched in β-sheet secondary structure. Dimers of this form would fit to the electron density previously found in single protofilaments of insulin. Further work which could be undertaken into this system could be an IM-MS analysis of whether the conformations of insulin oligomers change over time, particularly if there are changes in the multiple conformational families which are present in the dimeric to hexameric species.

The aggregation of Aβ peptides Aβ(1-40) and Aβ(1-42) and their interactions with aggregation inhibitor RI-OR2 were characterised by MS and IM-MS in Chapter Five. Monomeric species dominated the mass spectra in both negative and positive ionisation modes for all three peptides. RI-OR2 is designed to discourage self

[†] Professor Tom Rizzo's website http://lcpm.epfl.ch/page-53589-en.html

aggregation and MS indicated that this strategy has been successful as only a very small population of dimers and trimers were observed. Differences are observed in the oligomeric species populations present for Aβ(1-40) and Aβ(1-42), with higher order oligomers present for both Aβ peptides in positive ionisation mode. The largest Aβ(1-42) oligomers observed were dimers in negative ionisation mode and tetramers in positive ionisation mode. Higher order oligomers were discovered for Aβ(1-40); oligomers of tetrameric order in negative ionisation mode and hexameric order in positive ionisation mode.

When Aβ peptides were mixed with aggregation inhibitor RI-OR2 complexes between the two peptides were observed in small quantities. Various stoichiometries of complexes between the Aβ peptide and RI-OR2 were present, with 1:1 complexes generally the most populated. Additional stoichiometries detected were Aβ:2RI-OR2 and 2Aβ:RI-OR2. The use of IM-MS to probe the conformational landscape found that complexes of Aβ and RI-OR2 predominantly possessed substantially more compact CCSs than the corresponding Aβ species of the same charge state. This compaction of an elongated Aβ species could be the mechanism by which RI-OR2 inhibits Aβ oligomer formation. However two complexes of Aβ(1-40) and RI-OR2 with the highest charge states were more extended than the corresponding Aβ(1-40) monomeric charge state CCSs, possibly explaining why RI-OR2 is a less effective inhibitor of Aβ(1-40) aggregation. Further investigations, both experimentally and utilising molecular modelling, could be fruitful in characterising interactions between RI-OR2 and Aβ peptides and determining the interaction loci, providing insight into the mode of action of this potential therapeutic. A detailed time course of Aβ peptide aggregation could also give valuable clues to how the aggregation processes of Aβ(1-40) and Aβ(1-42) progress and differ.

The use of other experimental techniques to complement those employed in this thesis would be advantageous. In particular, concurrent fluorescence and EM time course studies, conducted in parallel to MS and IM-MS experiments, would allow direct comparison of temporal oligomer evolution and fibrillogenesis. Additionally CD experiments could be employed to monitor changes in the secondary structural

elements present in the aggregating solutions, although these measurements give population averages and not detail of individual oligomeric species. SEC and AUC, although of lower resolution than MS, would allow a quantitative comparison of the oligomeric species present in solution to be made with the relative abundances of the oligomeric ions observed in the gas phase by MS and IM-MS.

To conclude, the analytical techniques of MS and IM-MS, combined with molecular modelling, have proved to be powerful tools in elucidating details of the oligomeric species present in protein and peptide aggregation. The investigation of more aggregating systems would be rewarding. In addition to *in vitro* studies, the speed and sensitivity of these analytical techniques recommends their potential use in clinical diagnostic tests.

Appendices

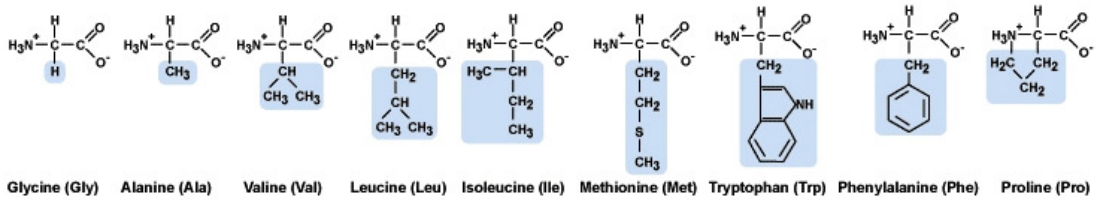
Appendix 1: Amino Acid Abbreviations

Amino Acid	Three Letter Code	One Letter Code	Monoisotopic Mass	Average Mass	Gas Phase Basicity * kcal/mol
Glycine	Gly	G	54.02146	57.05	202.7
Alanine	Ala	A	71.03711	71.08	206.4
Serine	Ser	S	87.03203	87.08	207.6
Proline	Pro	P	97.05276	97.12	214.3
Valine	Val	V	99.06841	99.07	208.7
Threonine	Thr	T	101.0477	101.1	211.7
Cysteine	Cys	C	103.0092	103.1	206.2
Leucine	Leu	L	113.0841	113.2	209.6
Isoleucine	Ile	I	113.0841	113.2	210.8
Asparagine	Asn	N	114.0429	114.1	212.8
Aspartic acid	Asp	D	115.0269	115.1	208.6
Glutamic acid	Glu	E	128.0586	128.1	215.6
Lysine	Lys	K	128.095	128.2	221.8
Glutamine	Gln	Q	129.0496	129.1	214.2
Methionine	Met	M	131.0405	131.2	213.3
Histidine	His	H	137.0589	137.1	223.7
Phenylalanine	Phe	F	147.0684	147.2	212.1
Arginine	Arg	R	156.1011	156.2	237.0
Tyrosine	Tyr	Y	163.0633	163.2	213.1
Tryptophan	Trp	W	186.0793	186.2	216.1

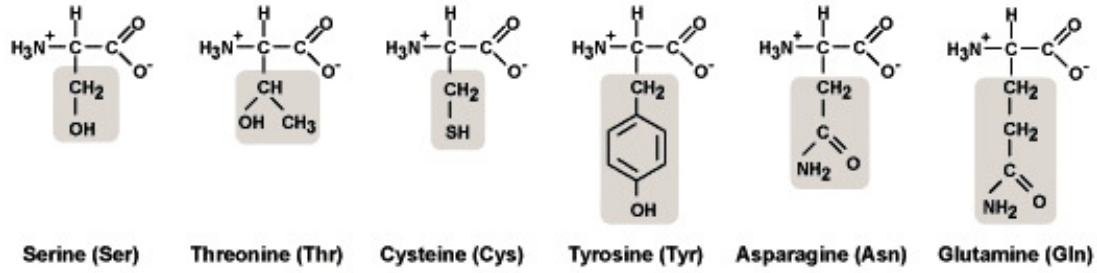
* Harrison, A. G. The gas-phase basicities and proton affinities of amino acids and peptides. *Mass Spectrom. Rev.* 16, 201-217 (1997).

Appendix 2: Amino Acid R Groups

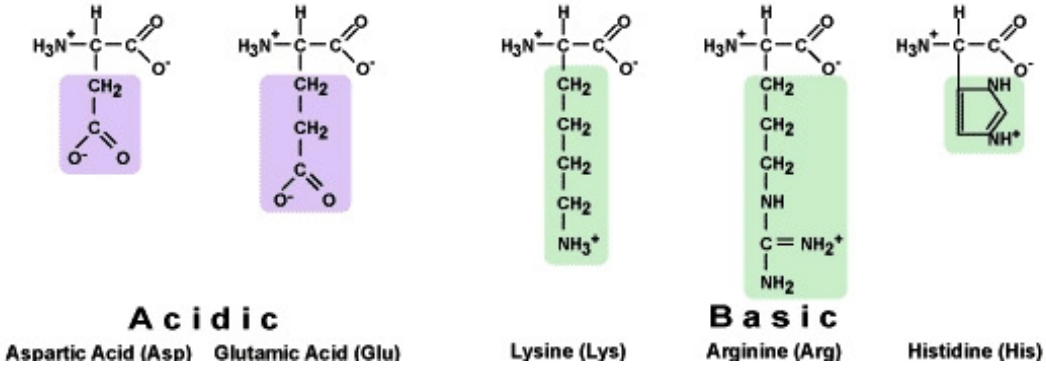
Non Polar Amino Acids



Polar Amino Acids



Electrically Charged Amino Acids



Figures from the website of the Department of Biology, Penn State University https://wikispaces.psu.edu/display/bio110/Carbon+and+Life

Appendix 3: TTR (105-115) CCSs

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[M+H]^{+}$	256.1	11.1
$[M+2H]^{2+}$	272.7	20.0
$[2M+2H]^{2+}$	380.2	24.5
$[3M+2H]^{2+}$	508.5	71.6
$[4M+2H]^{2+}$	443.1	17.8
$[4M+3H]^{3+}$	662.7	57.6
$[5M+3H]^{3+}$	665.7	51.5
$[5M+4H]^{4+}$	846.5	42.7
$[6M+4H]^{4+}$	848.3	85.1
$[6M+5H]^{5+}$	1013.5	75.9
$[7M+4H]^{4+}$	853.8	33.0
$[7M+5H]^{5+}$	1038.6	63.2
$[7M+6H]^{6+}$	1210.8	54.2
$[8M+5H]^{5+}$	1096.5	78.4
$[9M+5H]^{5+}$	1120.6	77.3
$[9M+7H]^{7+}$	1458.8	82.9
$[10M+6H]^{6+}$	1132.1	131.7
$[10M+7H]^{7+}$	1436.7	72.9
$[10M+9H]^{9+}$	1775.3	82.3
$[11M+6H]^{6+}$	1328.0	63.8
$[11M+7H]^{7+}$	1478.4	61.5
$[11M+10H]^{10+}$	1976.3	166.3
$[12M+7H]^{7+}$	1477.1	11.4
$[13M+12H]^{12+}$	2335.5	115.3

Table A1 | Rotationally averaged CCSs for all oligomeric species observable at t = 0. The CCSs show the mean and standard deviation of five repeated measurements.

Species	Time / hours									
	0		2		4		6		8	
	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation
$[M+H]^+$	264.8	10.0	259.5	15.3	269.0	3.8	266.3	6.3	264.7	14.6
$[M+2H]^{2+}$	263.1	4.0	263.0	5.3	263.0	4.7	260.2	2.7	259.4	0.6
$[2M+2H]^{2+}$	378.7	10.5	387.5	17.5	382.4	6.3	387.5	5.6	386.6	2.9
$[3M+2H]^{2+}$	508.9	101.5	505.4	74.4	527.6	106.3	513.8	117.0	524.7	94.8
$[4M+3H]^{3+}$	661.3	71.5	684.9	57.5	713.1	50.2	681.1	105.4	721.2	27.7
$[4M+2H]^{2+}$	445.0	18.7	472.1	-	480.7	1.3	497.0	20.1	483.5	26.2
$[5M+3H]^{3+}$	655.8	68.3	676.0	96.3	734.2	119.2	760.2	148.1	700.8	104.7
$[5M+4H]^{4+}$	846.5	42.6	850.9	40.2	862.0	40.4	937.2	163.4	875.0	41.2
$[6M+4H]^{4+}$	853.8	85.7	909.7	24.6	966.7	1.2	958.0	2.8	951.1	26.6
$[6M+5H]^{5+}$	997.9	86.3	997.2	97.5	1035.7	98.0	1167.0	239.2	1094.9	141.8
$[7M+4H]^{4+}$	860.9	48.6	1032.8	-	991.7	10.3	1036.2	22.6	989.8	69.9
$[7M+5H]^{5+}$	1027.5	65.9	1117.0	140.9	1105.5	142.4	1164.7	198.5	1088.6	126.5
$[7M+6H]^{6+}$	1224.1	69.4	1231.8	28.0	1294.7	131.2	1302.8	152.2	1310.0	184.6
$[8M+5H]^{5+}$	1071.0	56.6	1103.4	52.1	1138.5	61.9	1119.7	118.1	1128.0	77.0
$[9M+5H]^{5+}$	1111.3	46.8	1193.1	-	1196.5	3.5	1238.8	32.4	1201.1	71.8
$[10M+6H]^{6+}$	1135.4	137.4	1340.3	-	1368.7	18.5	1334.9	40.2	1261.6	89.1

Species	Time / hours									
	0		2		4		6		8	
	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation	CCS / Å^2	Standard Deviation
$[10M+9H]^{9+}$	1820.7	66.0	1811.2	64.9	1727.4	201.5	1924.7	334.1	1814.2	215.2
$[11M+6H]^{6+}$	1297.1	58.4	1408.5	-	1409.6	6.2	1441.1	44.0	1424.1	94.9
$[11M+10H]^{10+}$	1959.9	104.9	1919.5	73.9	2053.1	250.0	2082.9	261.5	2101.8	186.1

Table A2 | Rotationally averaged CCSs for all oligomeric species observable throughout an eight hour time course. The CCSs show the mean and standard deviation of five repeated measurements.

Appendix 4: Insulin CCSs

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[M+6H]^{6+}$	978.9	22.2
$[M+5H]^{5+}$	912.3	5.6
$[M+4H]^{4+}$	805.7	4.3
$[M+3H]^{3+}$	704.5	24.1
$[2M+8H]^{8+}$	1611.1	8.5
$[2M+6H]^{6+}$	1095.6	20.1
$[2M+5H]^{5+}$	1200.9	9.0
$[2M+4H]^{4+}$	961.3	17.0
$[3M+7H]^{7+}$	1682.1	14.6
$[4M+10H]^{10+}$	1907.1	84.5
$[4M+ 9H]^{9+}$	2160.4	102.6
$[4M+7H]^{7+}$	1754.2	40.7
$[5M+9H]^{9+}$	2265.7	35.0
$[5M+8H]^{8+}$	1982.8	57.4
$[5M+7H]^{7+}$	1680.3	8.3
$[6M+11H]^{11+}$	2695.9	86.1
$[6M+10H]^{10+}$	2465.2	41.2
$[6M+9H]^{9+}$	2205.2	26.2
$[6M+8H]^{8+}$	1888.6	62.3
$[7M+13H]^{13+}$	3113.1	64.7
$[7M+12H]^{12+}$	2938.0	66.6
$[7M+11H]^{11+}$	2791.2	68.2
$[7M+10H]^{10+}$	2399.6	82.6
$[8M+15H]^{15+}$	3608.7	102.6
$[8M+13H]^{13+}$	3198.7	60.7
$[8M+11H]^{11+}$	2575.3	63.6
$[9M+14H]^{14+}$	3476.1	29.5
$[9M+13H]^{13+}$	3146.3	62.0
$[11M+17H]^{17+}$	4134.9	46.7
$[12M+17H]^{17+}$	4040.2	36.3

Table A3 | Rotationally averaged CCSs for multimeric species observable at t = 0. The CCSs show the mean and standard deviation of three repeated measurements.

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[2M+7H]^{7+}$	1217.0	32.5
$[2M+7H]^{7+}$	1323.7	49.3
$[2M+7H]^{7+}$	1563.9	9.6
$[2M+7H]^{7+}$	1701.0	*
$[2M+9H]^{9+}$	1678.9	179.1
$[2M+9H]^{9+}$	2305.3	250.5
$[2M+9H]^{9+}$	2565.2	71.6
$[3M+8H]^{8+}$	1487.6	22.0
$[3M+8H]^{8+}$	1621.9	97.7
$[3M+8H]^{8+}$	1764.3	48.1
$[3M+8H]^{8+}$	1948.3	*
$[3M+10H]^{10+}$	1722.9	35.4
$[3M+10H]^{10+}$	2560.9	136.6
$[3M+10H]^{10+}$	1609.2	333.0
$[4M+11H]^{11+}$	1954.2	47.9
$[4M+11H]^{11+}$	2363.6	107.3
$[4M+11H]^{11+}$	2731.0	7.2
$[4M+11H]^{11+}$	3034.8	*
$[5M+11H]^{11+}$	2259.4	64.3
$[5M+11H]^{11+}$	2516.9	30.5
$[5M+11H]^{11+}$	2877.9	*
$[5M+12H]^{12+}$	2249.5	49.6
$[5M+12H]^{12+}$	2622.0	4.7
$[5M+12H]^{12+}$	2911.3	107.5
$[5M+12H]^{12+}$	3418.7	*
$[6M+13H]^{13+}$	2207.9	148.5
$[6M+13H]^{13+}$	2663.4	108.7
$[6M+13H]^{13+}$	2997.8	80.2
$[6M+13H]^{13+}$	3246.5	*
$[6M+13H]^{13+}$	3673.1	*

*Table A4 | Rotationally averaged CCSs for multimeric species present with several conformations observable at t = 0. The CCSs show the mean and standard deviation of three repeated measurements. * represents species which were only identified in one repeat and therefore no standard deviation could be calculated.*

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[5M+7H]^{7+}$	1263.3	21.0
$[5M+8H]^{8+}$	1490.5	109.6
$[6M+8H]^{8+}$	1463.4	17.0
$[6M+9H]^{9+}$	1641.2	17.3
$[6M+11H]^{11+}$	1971.2	148.8
$[7M+10H]^{10+}$	1803.3	74.7
$[7M+11H]^{11+}$	1996.3	48.3
$[7M+12H]^{12+}$	2183.0	46.8
$[7M+13H]^{13+}$	2291.8	122.6
$[8M+11H]^{11+}$	1964.7	79.8
$[8M+13H]^{13+}$	2399.0	9.3
$[8M+15H]^{15+}$	2763.5	156.3
$[9M+13H]^{13+}$	2306.6	41.5
$[9M+14H]^{14+}$	2537.0	23.2
$[11M+17H]^{17+}$	3100.3	59.9
$[12M+17H]^{17+}$	2992.3	105.4

Table A5 | If the earlier arriving ATD peaks are assigned as compact species the above rotationally averaged CCSs result. The CCSs show the mean and standard deviation of three repeated measurements.

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[10M+14H]^{14+}$	2526.6	41.9
$[10M+16H]^{16+}$	2981.0	219.3
$[12M+16H]^{16+}$	2926.7	33.9
$[12M+18H]^{18+}$	3282.4	34.6
$[12M+22H]^{22+}$	3942.4	297.5
$[14M+20H]^{20+}$	3606.6	149.3
$[14M+22H]^{22+}$	3992.5	96.5
$[14M+24H]^{24+}$	4366.0	93.6
$[14M+26H]^{26+}$	4583.6	245.1
$[16M+22H]^{22+}$	3929.4	159.5
$[16M+26H]^{26+}$	4798.1	18.5
$[16M+30H]^{30+}$	5527.1	312.6
$[18M+26H]^{26+}$	4613.2	83.1
$[18M+28H]^{28+}$	5073.9	46.4
$[22M+34H]^{34+}$	6200.6	119.8
$[24M+34H]^{34+}$	5984.7	210.9

Table A6 | If the earlier arriving ATD peaks are assigned as coincident m/z oligomers the above rotationally averaged CCSs result. The CCSs show the mean and standard deviation of three repeated measurements.

Appendix 5: Aβ and RI-OR2 CCSs

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[(RI\text{-}OR2)+3H]^{3+}$	302.9	0.8
$[(RI\text{-}OR2)+2H]^{2+}$	246.0	2.3

Table A7 | Rotationally averaged CCSs of RI-OR2. The CCSs show the mean and standard deviation of three repeated measurements.

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[Aβ40+2H]^{2+}$	333.0	3.9
$[Aβ40+3H]^{3+}$	567.8	8.8
$[Aβ40+4H]^{4+}$	650.8	6.9
$[Aβ40+5H]^{5+}$	725.3	13.8
$[Aβ40+6H]^{6+}$	767.2	12.3
$[2Aβ40+4H]^{4+}$	665.9	7.9
$[2Aβ40+5H]^{5+}$	912.2	31.5
$[2Aβ40+6H]^{6+}$	999.5	86.1
$[2Aβ40+7H]^{7+}$	1115.3	134.9
$[2Aβ40+9H]^{9+}$	1274.2	34.9
$[3Aβ40+5H]^{5+}$	1007.6	51.4
$[3Aβ40+7H]^{7+}$	1217.5	41.7
$[4Aβ40+7H]^{7+}$	1249.2	11.4
$[4Aβ40+9H]^{9+}$	1528.8	40.7
$[4Aβ40+14H]^{14+}$	2033.4	126.5
$[5Aβ40+13H]^{13+}$	2027.7	58.5
$[6Aβ40+10H]^{10+}$	1938.4	38.9

Table A8 | Rotationally averaged CCSs of Aβ(1-40). The CCSs show the mean and standard deviation of three repeated measurements.

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[Aβ42+2H]^{2+}$	375.4	23.3
$[Aβ42+3H]^{3+}$	552.9	48.1
$[Aβ42+4H]^{4+}$	646.5	14.3
$[Aβ42+5H]^{5+}$	742.1	10.7
$[Aβ42+6H]^{6+}$	757.4	9.7
$[Aβ42+7H]^{7+}$	771.9	72.2
$[2Aβ42+4H]^{4+}$	750.6	46.6
$[2Aβ42+5H]^{5+}$	860.2	88.6
$[2Aβ42+6H]^{6+}$	985.6	114.0
$[2Aβ42+7H]^{7+}$	1022.3	25.4
$[3Aβ42+6H]^{6+}$	1125.9	69.9
$[4Aβ42+9H]^{9+}$	1623.7	43.2

Table A9 | Rotationally averaged CCSs of Aβ(1-42). The CCSs show the mean and standard deviation of three repeated measurements.

	CCS / Å²	
Species	**Average**	**Standard Deviation**
$[Aβ40+(RI-OR2)+3H]^{3+}$	497.7	21.0
$[Aβ40+(RI-OR2)+4H]^{4+}$	582.8	21.8
$[Aβ40+(RI-OR2)+5H]^{5+}$	780.0	17.5
$[Aβ40+(RI-OR2)+6H]^{6+}$	819.6	52.5
$[Aβ40+(2RI-OR2)+3H]^{3+}$	507.2	24.3
$[Aβ40+(2RI-OR2)+4H]^{4+}$	635.8	30.2
$[2Aβ40+(RI-OR2)+5H]^{5+}$	837.7	32.1
$[Aβ42+(RI-OR2)+3H]^{3+}$	528.0	10.3
$[Aβ42+(RI-OR2)+4H]^{4+}$	597.2	30.0
$[Aβ42+(2RI-OR2)+4H]^{4+}$	645.3	23.0

Table A10 | Rotationally averaged CCSs of complexes between Aβ and RI-OR2. The CCSs show the mean and standard deviation of three repeated measurements.

Appendix 6: Publications, Conference Papers and Articles

Cole, H.L., Kalapothakis, J.M.D., Bennett, G., Barran, P.E., & MacPhee, C.E., Characterizing Early Aggregates Formed by an Amyloidogenic Peptide by Mass Spectrometry. *Angew. Chem.-Int. Edit.* **2010** 49, 9448-9451.

Barran, P.E., Cole, H.L., Goldup, S.M., Leigh, D.A., McGonigal, P.R., Symes, M.D., Wu, J., Zengerle, M., Active-Metal Template Synthesis of a Molecular Trefoil Knot. *Angew. Chem.-Int. Edit.* **2011** 50, 12280-12284.

Cole, H.L., Porrini, M., MacPhee, C.E., Barran, P.E., Insights into the early stages of fibrillogenesis of insulin using mass spectrometry **2011** 8th EBSA European Biophysics Congress. European Biophysics Journal with Biophysics Letters Volume: 40 Supplement: 1 Pages: 88-88

Cole, H.L., Kalapothakis, J.M.D., MacPhee, C.E., Barran, P.E., The Use of Ion Mobility Mass Spectrometry to Study Pre-fibrillar Aggregates. *Mass Matters* **2011** 65